A COMPUTATIONAL STUDY OF THE ROLE OF HYDRATION IN THE ASSEMBLY OF COLLAGEN AND OTHER BIOFILAMENTS

A Dissertation

by

KRISHNAKUMAR MAYURAM RAVIKUMAR

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2011

Major Subject: Biomedical Engineering

A COMPUTATIONAL STUDY OF THE ROLE OF HYDRATION IN THE ASSEMBLY OF COLLAGEN AND OTHER BIOFILAMENTS

A Dissertation

by

KRISHNAKUMAR MAYURAM RAVIKUMAR

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Wonmuk Hwang
Committee Members,	Alvin T. Yeh
	Roland R. Kaunas
	Mariappan Muthuchamy
Head of Department,	Gerard L. Coté

August 2011

Major Subject: Biomedical Engineering

ABSTRACT

A Computational Study of the Role of Hydration in the Assembly of Collagen and Other Biofilaments. (August 2011)

Krishnakumar Mayuram Ravikumar, B.E., University of Madras;

M.S., Texas A&M University

Chair of Advisory Committee: Dr. Wonmuk Hwang

Hydration is known to be crucial in biomolecular interactions including ligand binding and self-assembly. In our earlier studies we have shown the key role of water in stabilizing the specific parts of the collagen triple helix depending on the imino acid content. We further showed that the primary hydration shell around collagen could act as a lubricating layer aiding in collagen assembly. But key details on the structure and dynamics of water near protein surfaces and its role in protein-protein interactions remain unclear. In the current study we have developed a novel method to analyze hydration maps around peptides at 1-Å resolution around three self-assembling filament systems with known structures, that respectively have hydrated (collagen), dry non-polar and dry polar (amyloid) interfaces. Using computer simulations, we calculate local hydration maps and hydration forces. We find that the primary hydration shells are formed all over the surface, regardless of the types of the underlying amino acids. The weakly oscillating hydration force arises from coalescence and depletion of hydration shells as two filaments approach, whereas local water diffusion, orientation, or hydrogen bonding events have no direct effect. Hydration forces between hydrated, polar, and non-polar interfaces differ in the amplitude and phase of the oscillation relative to the equilibrium surface separation. Therefore, water-mediated interactions between these protein surfaces ranging in character from 'hydrophobic' to 'hydrophilic,' have a common molecular origin based on the robustly formed hy-

dration shells, which is likely applicable to a broad range of biomolecular assemblies whose interfacial geometry is similar in length scale to those of the present study.

In a related study through simulations we show that the rate of tissue optical clearing by chemical agents correlated with the preferential formation of hydrogen bond bridges between agent and collagen. Hydrogen bond bridge formation disrupts the collagen hydration layer and facilitates replacement by a chemical agent to destabilize the tertiary structure of collagens thereby reducing light scattering. This study suggests that the clearing abiility of an alcohol not only depends on its molecular size, but also on the position of hydroxyl groups on its backbone.

ACKNOWLEDGMENTS

First, I thank my advisor Wonmuk Hwang for his continued guidance and support all through my PhD. He was instrumental in inculcating my interest in statistical mechanics, biomolecular systems, and water. Consciously and sometimes unconsciously he has taught be great values in life including hard work, patience, and determination to help me perform beyond my capabilities. He has always been open to discussion and I have always loved the freedom to barge into his office anytime to discuss. In spite of his busy schedule he has always made time for me and I am very grateful and thankful for that. During frustrating times in research he was always there for me helping with excellent ideas and suggestions and not letting me get bogged down. I am also grateful for the opportunities he gave me to travel for conferences and meet people to learn from various other research groups. I have always looked up to him as a role model to learn from. More than anything I have become a better human being working under him.

I would like to thank all Hwang group lab members (Kaushik, Jiyong, John, Weewen, Francesca, Jinseon, Xiaojing, Andrew, Tyler, Esma, Amol) who have always helped me in discussions, poster preparations, and presentations. Their feedback was crucial in many ways to improve the quality of my research. I would surely cherish their friendship and collaborations for life. I thank Jiyong (visiting student at our lab) for teaching me basics of Linux and molecular dynamics simulations. I am thankful for his patience to put up with all my rudimentary questions on Linux and CHARMM. I thank Tyler for running preliminary test runs on β sheets. I am also grateful to Kaushik and Jiyong for many stimulating discussions we have had about simulation techniques, and biomolecules in general. I cannot thank Kaushik enough since he has always been there to help me with almost everything in the lab (fixing

computers, install Linux, or research). Such great camaraderie in the lab made life so much easier.

I would also like to thank Alvin T. Yeh and Jason Hirshburg who helped me with interesting ideas to analyze collagen-alcohol simulations. Their discussions were very helpful. I thank Jay D. Humphrey for insightful discussions about collagen conformational fluctuations and their implications. I also thank Alvin T. Yeh and Jay D. Humphrey for critically reading and commenting on our papers. I thank my other committee members, Mariappan Muthuchamy and Roland R. Kaunas for critical comments for my research and helpful suggestions to improve my research.

My life at Texas A&M was enjoyable to a great extent because of the many friends, roommates, and music lovers. I have thoroughly enjoyed music practice sessions and concerts at SPICMACAY (Society for Promotion of Music and Culture Amongst Youth). I would like to thank all SPICMACAY members, its advisors Shankar Bhattacharyya (Electrical Engineering), Mysore Mohan (Chemistry) and music connoisseurs Sivakumar N (Mathematics), Krishna Narayanan (Electrical Engineering), and Swaroop Darbha (Mechanical Engineering) for their friendly support.

Lastly, I would like to thank my family, especially my mother and sister for continuously supporting me in all my pursuits. Without their support I couldn't have done most things in life.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION: MOTIVATION AND PREVIOUS STUDIES ON HYDRATION OF COLLAGEN*

A. Introduction

Currently 28 types of collagen molecules are known, out of which more than 90% are fibril forming and structurally similar types I,II, and III [1, 2, 3]. Collagen is the major component of extracellular matrix (ECM) and is commonly classified as a structural protein since it provides tensile strength to tissues. Unlike traditional engineering materials, collagens (≈300 nm in length) are dynamic building blocks designed to self-assemble in an orderly fashion to form hierarchical structures - microfibrils, fibrils, fibers, and matrix. Cells can regulate collagen assembly, tissue growth, and turnover using MMPs, which can cleave collagen. Irregular assembly or turnover of collagen will lead to several pathological conditions [4]. The complex yet robust design of collagen assembly is thus important from both bioengineering and biomedical point of view, but understanding collagen assembly poses an interesting challenge as it includes multiple length and time scales.

Monomeric collagen molecule has a rope-like structure, where three left-handed helical poly-proline-II type α chains are wound to form a right-handed triple helical structure [2]. X-ray crystal structures of collagen mimetic peptides have provided

---

This dissertation follows the style of *IEEE Transactions on Automatic Control.*

*Portions of this chapter are reprinted with permission from K. M. Ravikumar, J. D. Humphery, and W. Hwang, "Spontaneous unwinding of a labile domain in a collagen triple helix," *J Mech Mater Struct*, vol. 2, pp. 999–1010, 2007. Copyright 2007 by Journal of Mechanics of Materials and Structures and from K. M. Ravikumar and W. Hwang, "Region-specific role of water in collagen unwinding and assembly." *Proteins*, vol. 72, no. 4, pp. 1320–1332, 2008. Copyright 2008 by Wiley-Liss Inc.

vital insights into the molecular structure and function of the collagen monomers. Each collagen alpha chain has a -GXY- repeating sequence, where G is glycine and X,Y are other amino acids. Regular backbone hydrogen bonds between α chains are important in stabilizing the molecule [5, 6]. Imino-acids PRO and HYP (Proline and Hydroxyproline) impart stability to the molecules as regions with high imino acid content (imino-rich) in collagen form a tightly wound helical structure, while imino-acid deficient (imino-poor) regions form a more loosely wound helix [6]. Such local differences in the helical twist of the tertiary helical structure of collagen have important functional implications. It is known that the cleavage site on collagen is loosely wound to aid cleavage enzymes which can only cut individual α chains but not all the three α chains of collagen together [7]. Local unwinding thus provides a mechanism by which external load on collagen can affect its cleavage and tissue turnover. Individual collagen molecules are unstable at body temperature and have to self-assemble into higher order structures once secreted [8]. Unassembled collagen monomers are unstable and can be easily cleaved by the cleavage enzymes.

Unwinding and stability of the monomers have been studied using collagen mimetic peptides through calorimetry [9] which showed increased stabilizing effect of HYP content. It was thought that HYP confers higher stability through hydrogen bonding with water, but the mechanism of HYP stabilization has been shown to be different [10]. Raines et.al. showed that inductive effect of the hydroxyl group confers a tighter helical twist to the collagen molecules. But, water does play a stabilizing role [11] in the imino-poor regions, where water dynamically forms hydrogen bond bridges with the collagen backbone. Owing to the lose packing of the imino-poor regions, water oxygen atoms are much closer to the backbone compared to the imino-rich regions and forms hydrogen bonds with the backbone amide hydrogen (-N-H) atoms of collagen. At a higher length scale, water has an important role in the

collagen assembly. Water molecules orient themselves in an orderly fashion forming hydrogen bonds with the backbone leading to a clear definition of the first hydration layer, at the distance where two collagen molecules self-assemble [11]. Thus water acts as a lubricant in collagen self-assembly. Our geometric picture of hydration gives insight into the hydration structure and its role in self-assembly, but knowledge of water dynamics in the hydration layer and its effects on hydration forces are still missing. Water molecules will lose their translational and rotational mobility near protein surfaces [12]. The nature and effect of this loss on collagen self-assembly is an open question.

This leads us to the wider and general problem of protein hydration. It is believed that hydration plays an important role in imparting a delicate balance between rigidity and flexibility of protein structures. The rigidity is necessary for substrate specific interactions, while enough flexibility helps the protein to adapt its shape while binding and self-assembly [13]. But how water decorates folded protein structures and how protein and water dynamics affect each other remains unclear [13]. Protein surfaces often have polar and non-polar sidechains side-by-side, and the effect of surface polarity on local water structure is also unclear. Our previous studies provided some insights in this regard, but further studies with higher spatial resolution are required to clarify these questions.

B. Previous Results

1. Water Bridges and Local Unwinding of Collagen

Local microunfolding of the collagen triple helix was speculated in previous studies [15, 16] and is known to have important functional implications. For example, MMPs cannot cleave tightly wound triple helical collagen as its catalytic cleavage domain

can accommodate only a single collagen α chain [7]. To understand the conformational fluctuations of individual molecules, we studied the unwinding behavior of the collagen mimetic peptide 1BKV taken from Protein Data Bank (PDB), through Molecular Dynamics (MD) simulations using Charmm-22 all-atom force field [17]. Peptide 1BKV has a biologically relevant sequence (domain II in Fig. 1) from type III collagen, which is known to be important in collagen cleavage [6, 15]. We used a method to monitor the torsional motion of collagen at amino-acid level resolution. Briefly, we defined planes perpendicular to the cross-section of the peptide by joining the Cα atoms of adjacent α chains. We then defined orthonormal basis (triad) vectors at each cross section and monitored the Euler angle changes between any triad to the first triad (Fig. 2) [18] to get a *torsional map* of the peptide. Local unwinding is reflected as an inflexion in the torsional map. At 273 K there was no unwinding within the simulation time, but at 300 K and 330 K the peptide unwound in the imino-poor (domain II) region (Fig. 2). Unwinding was triggered at the Gly-Ile bond region probably owing to Ile's high β-sheet propensity.

We also investigated the role of water in unwinding. In the imino-poor domain water formed hydrogen bond bridges (lifetimes in the order of few picoseconds) with the backbone of the peptide. The degree of unwinding correlated with the number of water hydrogen bond bridges suggesting the role of water in preventing unwinding of the imino-poor region of collagen [11]. Further to confirm the role of water, we either enforced or prevented water bridges which resulted respectively in stabilizing or destabilizing the peptide. Monitoring local microunfolding and conformational fluctuations at high resolution showed the stabilizing role of water bridges in collagen unwinding.

2. Hydration Dynamics

We analyzed the hydration structure of water around the peptide by measuring the distance of water oxygen atoms from the protein backbone and defining the radial distribution function (RDF) of water around the protein (Fig. 3). Note that the backbone atoms are ≈3.0 - 4.0 Å from the helical axis of collagen. The RDF shows a density peak, which defines the first hydration shell, at ≈2.8 Å from the collagen backbone. This peak is followed by the *void* region with lower water oxygen density at ≈3.25 Å (≈6.25-7.25 Å from the axis of collagen). The peaks and troughs in the RDF arise from the orientations of water atoms at the protein surface [11]. More importantly, the first hydration shell forms at the distance where two collagen molecules self-assemble (≈12.5-14.5 Å) as seen in x-ray crystal structures [11, 19]. These results suggest that the first hydration layer acts like a lubricating layer in collagen self-assembly. Our results show the dual role of water in collagen stability (Fig. 1) and assembly.

C. Limitations

The RDF is spatially averaged for a given radial distance from the collagen backbone atoms which is not enough to differentiate the local spatial variation in the density near different surface groups. The RDF does not give any information about the dynamics of water in the hydration layers, which is also important in most biological interactions including collagen self-assembly [20, 21]. Spatial variation in water properties around protein molecules are difficult to study through experiments and also in MD simulations as well, owing to the motion of protein during simulation. Earlier simulation studies on hydration of other proteins have shown loss in translational and rotational mobility of water, but they were short simulations on single molecule

systems and partitioned the hydration layers into larger discrete regions [22, 23] for convenience. Longer simulations, with finer spatial discretization and analysis on both single and self-assembled protein systems are required for deeper insights into hydration dynamics.

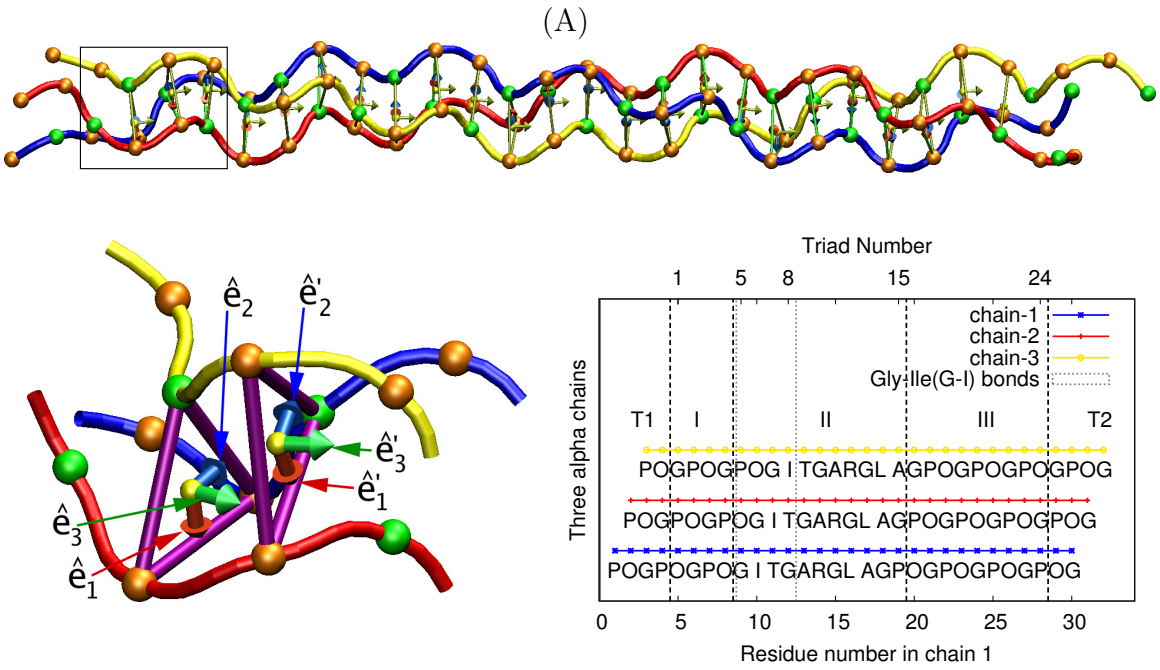

Fig. 1. Triads along the helix. The three α-chains are shown in blue (chain-1), red (chain-2), and yellow (chain-3). Spheres are C_α atoms of Gly (green) and X/Y (orange). (A) Overview of 1bkv with the 24 triads. (B) Magnified view of the box in (A). To show the face of the triangle, the molecule is rotated about the vertical axis of the paper plane. Two neighboring triads are denoted $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ and $\{\hat{e}'_1, \hat{e}'_2, \hat{e}'_3\}$. (C) Triad numbers versus amino acid sequence. The horizontal shift of the three chains reflects their staggered structure in the triple helix. Triads are grouped into domains I, II, and III (I,II - regular Gly-Pro-Hyp regions, II-Biologically relevant sequence from type-III Collagen). T1 and T2 are unconsidered regions to eliminate end effects. Triads 5 to 8, marked between gray dotted vertical lines, contain Gly-Ile bonds. All molecular images in this manuscript were rendered in VMD [14].

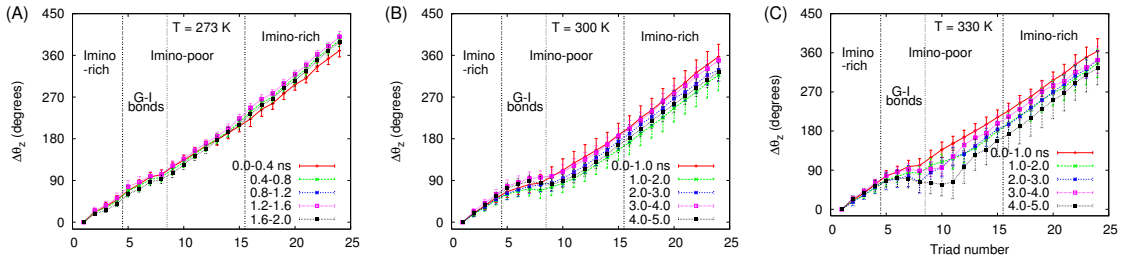

Fig. 2. Temperature-dependent unwinding. The local slope of each curve is directly proportional to the helicity at the corresponding triad. Each point corresponds to the helical angle difference from the first triad averaged over successive (A) 0.4-ns, (B),(C) 1-ns intervals (± standard deviation). Triads 5-8 contain Gly-Ile bonds, where unwinding is initiated.

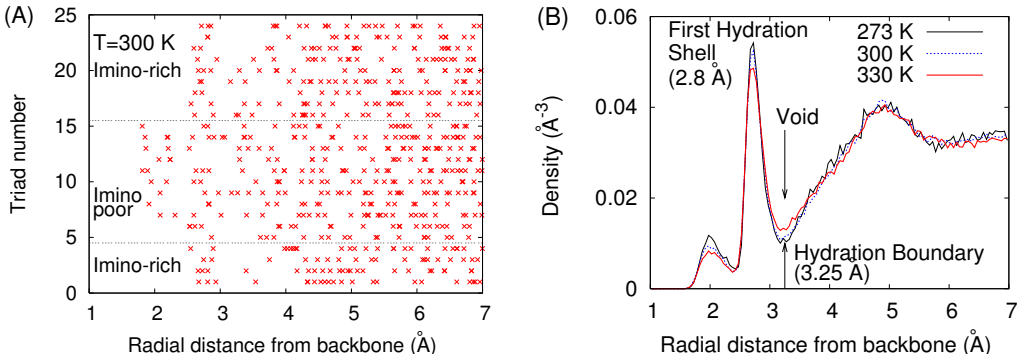

Fig. 3. Hydration structure around the collagen triple helix. (A) Snapshot of water oxygens (at 970 ps) at 300 K, Water oxygens are closer to the backbone in the imino-poor region compared to imino-rich regions. (B) Average radial distribution of water oxygen atoms over the simulation period (273 K: 2ns; 300,330 K: 5ns). The overall profile is preserved during the simulation, although the peak height decreases slightly as temperature increases.

CHAPTER II

ROLE OF HYDRATION FORCE IN THE SELF-ASSEMBLY OF COLLAGENS AND β-SHEET FILAMENTS*

A. Introduction

Water-mediated forces play crucial roles in biomolecular interactions and assemblies, yet understanding its physical basis still remains a challenge [28, 29, 30, 13]. Conventionally, these can be divided into hydrophobic and hydrophilic, where previous theoretical modeling and computer simulations provided much insight. Hydrophobic attraction between smooth or confined surfaces in close separations is known to arise from dewetting transition driven by the solvent fluctuation at the interface [31, 32, 33], while other effects such as polarization of water and solute, or nanobubble formation may be responsible for longer-range attraction [34, 35, 36, 37]. Between polar or charged surfaces that can form hydrogen bonds with water, the interaction can be either attractive, repulsive or oscillatory [38, 28]. Such *hydration force* is believed to be due to the ordering of water into solvation (hydration) shells around the solute surface [20, 38], where resistance to the removal of water from the surface is responsible for the repulsive force [39, 40]. The oscillatory behavior of the hydration force observed between macroscopically flat surfaces is due to the layering of hydration shells, which can smooth out to a monotonic profile when the surfaces are flexible [28, 20]. Another possible contributor to the hydration force is the influence of the surface on the orientational distribution of water [41].

---

*Portions of this chapter are reprinted with permission from K. M. Ravikumar, and W. Hwang, "Role of hydration shell in the self-assembly of collagens and amyloid steric zipper filaments." *J Am Chem Soc*, in press.

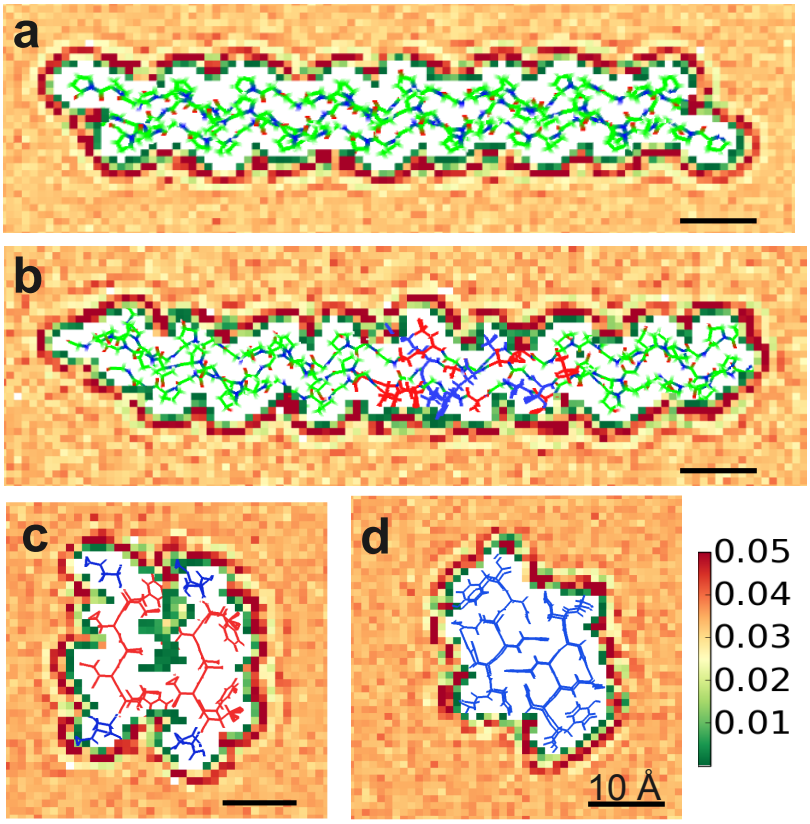

Fig. 4. Water density maps (in #/Å^3). (a,b) collagen in side view, and (c,d) amyloid β-sheet bilayers in axis view. Amino acids are colored blue (polar/charged), red (non-polar), and green (collagen Gly-Pro-(hydroxy)Pro triplets). The primary hydration shell and more faintly the secondary hydration shell form throughout the surface, irrespective of the types of the underlying amino acids. Protein Data Bank (PDB) IDs: (a) 2D3F [24], (b) 1BKV [6, 25], (c) 2KIB [26], and (d) 1YJP [27].

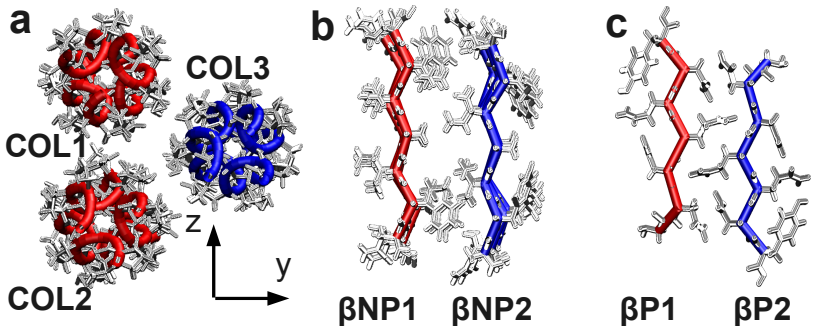

Fig. 5. Cross sectional view of the filaments used for force measurement. (a) Three 2D3F peptides. Individual collagen triple helices are named COL1–3. (b) 2KIB and (c) 1YJP. 'NP' and 'P' in β-sheet names stand for non-polar and polar, respectively. Filament axes are perpendicular to the page. COL3, βNP2, and βP2 are translated in the y-direction and forces on them during simulations are measured (see Methods). The anti-parallel β-sheet 2KIB forms a 'hetero zipper' [26], which is less tight compared to the interdigitation of the parallel β-sheet 1YJP [27].

At nanometer length scales, simulations provide an atomistic picture for water-mediated forces. However, most simulations used simplified surfaces such as plates, cylinders, spheres, or simplified protein structures [42, 39, 43, 44]. Others that have employed atomistic structures focused on hydrophobic attraction [45, 46, 47, 48]. It is unclear to what extent results for macroscopic or simpler systems can be extrapolated to biomolecular surfaces that are geometrically complex and contain various polar and non-polar groups side-by-side. For example, surface property changes from hydrophilic to hydrophobic as its roughness increases [49, 50]. Study of a model hydrophobic ligand-receptor complex even suggests that hydrophobic association, which is considered to be entropically driven, can be enthalpic in nature [51].

Here we study the hydration forces in three biomolecular filament systems that self-assemble with different types of interfaces (Fig. 4): Collagens remain hydrated after assembly [52, 6, 53], and the two β-sheet bilayer filaments form dry interfaces that respectively have non-polar and polar residues [26, 27]. The dry interface within the bilayer is also called a 'steric zipper' due to the geometric complementarity of the amino acid side chains (Fig. 5b,c) [27, 54]. We find that despite the presence of various amino acid side chains, hydration shells form in all cases, suggesting that the general tendency of a liquid to form high-density solvation shells near flat rigid surfaces ("hard-wall effect") [55, 56] apply for these filamentous protein surfaces. As a result, weakly oscillating hydration force arises regardless of the type of the surface considered, where quantitative details such as the magnitude and location of the hydration repulsion or attraction – the amplitude and phase of the hydration force – depend on surface polarity and geometry. The present results provide a unifying picture for water-mediated interactions, where the primary hydration shell plays a central role. Thus, at the length scale of surfaces studied here, the hydrophobic or hydrophilic nature of an interaction may not simply be an issue of the types of the

surface groups, but other factors such as surface geometry, complementarity, and flexibility may also play a role in the net interaction.

B. Methods

1. Peptides Used

We used x-ray structures of three collagen peptides, PDB 2D3F [24], 1BKV [6, 25], and 1A3I [19]. PDB 2D3F has the sequence $(PPG)_4$-POG-$(PPG)_4$ ($\approx$ 8-nm long; G: glycine, P: proline, and O: hydroxyproline). The GPP and GPO triplets are the most stable structural motifs in native collagen [11, 6]. We used a three-peptide 2D3F system for force measurements (Fig. 5a) and a single 2D3F to study the effect of protein diffusion on hydration maps. PDB 1BKV has a 9-residue region containing both non-polar and polar (including charged) residues in the middle. We used it to test hydration maps around the bulky polar and non-polar side chains in collagen triple helix. PDB 1A3I has three GPP units ($\approx$2-nm long), which is less than a third of the size of 2D3F or 1BKV. We used it to analyze the effect of different water models on hydration maps.

For β-sheet bilayers, we used PDB 2KIB [26] and 1YJP [27] (Fig. 4c,d). Unlike other x-ray structures used in our study, PDB 2KIB is a solid-state NMR structure, and we used the first structure among the ten in the PDB file. The amino acid sequence of 2KIB is NFGAILS, where F, A, I and L are non-polar. On the other hand, all residues of 1YJP (GNNQQNY) are polar. Each β-sheet forming a bilayer filament of 2KIB (1YJP) has 4 (5) peptides, about 1.66 nm (2 nm) in length.

2. Simulation Protocol

For MD simulation, we used the GROMACS simulation package [57] with all-atom CHARMM Param22 force field [58]. The peptides were solvated in an orthorhombic simulation box whose dimension ranges from $45 \times 45 \times 45$ Å^3 to $110 \times 55 \times 50$ Å^3, depending on the size of the peptides used. We used the SPC water model for the collagen simulations and the TIP3P model for those of β-sheet filaments. We also used the SPC/E water model to compare the calculated forces. Periodic boundary conditions were imposed. Additional force-field parameters for hydroxyproline were taken from an earlier study [59]. The system was energy minimized for 500 steps using the steepest descent method, heated from 0 K to target temperatures (273, 300, or 330 K) for 50 ps, and equilibrated at the respective temperatures for 30 ps, with all heavy atoms harmonically constrained to their original positions (spring constant k=10^4 kJ/(mol·nm^2)). The final production run was performed using the leap-frog integrator with a 2-fs time step. Hydrogen bonds were constrained using the LINCS algorithm [60]. Coordinates were saved every $\Delta t = 1$ ps. The non-bonded pair and image atom lists were updated every 20 fs. A 13-Å cutoff was used for non-bonded interaction energies and the particle mesh Ewald summation method [61] was used to calculate electrostatic interactions. Temperature and pressure were maintained using the velocity rescaling thermostat [62] and Berendesen pressure coupling [63]. Trajectories were stable during production runs with relative root-mean-square fluctuations of temperature and energy lesser than 1.2%.

In simulations of 1BKV that has a charge of $+3e$ ($e = 1.6 \times 10^{-19}$C), we added three Cl^- ions to neutralize the net charge. The production run for each of the 3-peptide 2D3F and single-peptide 1A3I simulations lasted 8 ns, while it was 4 ns for single-peptide 1BKV or 2D3F simulations. All β-sheet simulations lasted for 8 ns.

The total simulation time for the entire study was $\approx 1.4\mu$s.

3. Hydration Map

Within each 1-Å^3 unit cell of the simulation box, we calculated the following:

Density. If a water oxygen visits the cell n times, the density is $\rho = n/n_{tot}$, where n_{tot} is the total number of saved coordinate frames.

Translational diffusion coefficient. For frames where a water oxygen visits the cell, we calculated its mean-square displacement during Δt (= 1 ps), $\langle \Delta l^2 \rangle = \sum \Delta l^2 / n$. The diffusion coefficient is $D_t = \langle \Delta l^2 \rangle / 6\Delta t$. This definition gives known values of diffusion coefficients in the bulk water for the water models we tested (Fig. 6).

Orientational diffusion coefficient. We assign three unit vectors to a water molecule: $\mathbf{r}_1$, along the water dipole; $\mathbf{r}_2$, orthogonal to $\mathbf{r}_1$ in the plane containing water atoms; and $\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2$ (Fig. 7). Denoting the 2-dimensional angular displacement of $\mathbf{r}_i$ $(i = 1, 2, 3)$ by ϕ_i, the corresponding orientational diffusion coefficient is $D_{ri} = \langle \Delta\phi_i^2 \rangle / 4\Delta t$, with $\langle \Delta\phi_i^2 \rangle = \sum \Delta\phi_i^2 / n$.

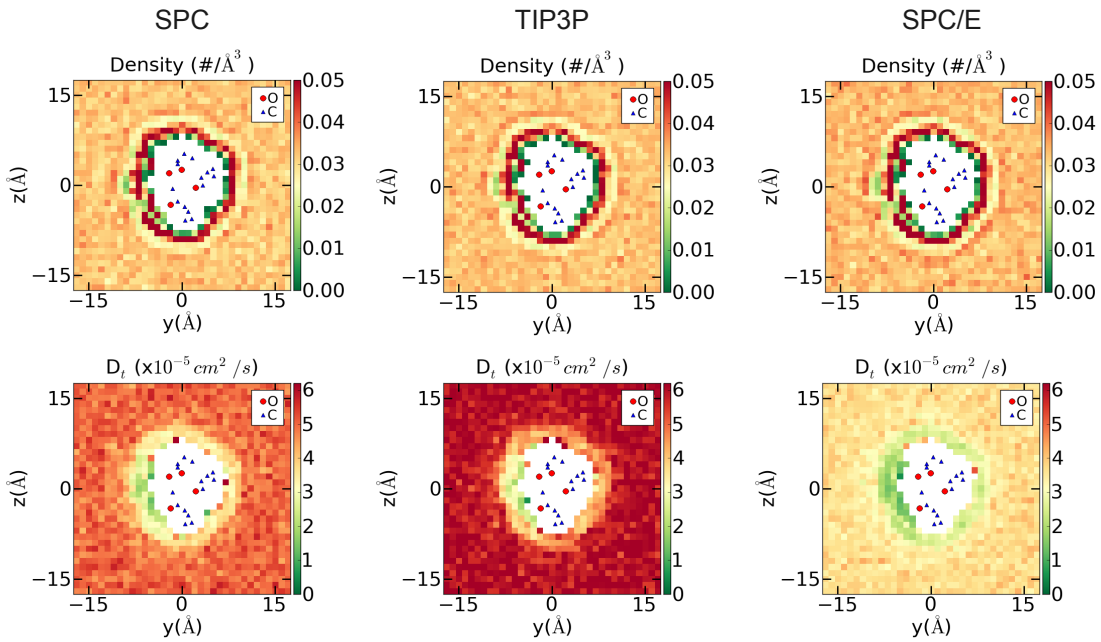

Fig. 6. Comparison between different water models at 300 K on a collagen peptide 1A3I. The density maps are nearly the same for all water models. While the diffusion coefficient of the SPC/E model in the bulk is closer to the experimental value compared to SPC or TIP3P [64], the diffusion coefficient maps are qualitatively similar among the three water models. Together with the model-independence of the measured force (Fig. 21 and Fig. 22), this further supports the idea that hydration force is generated by packing of water and insensitive to water dynamics.

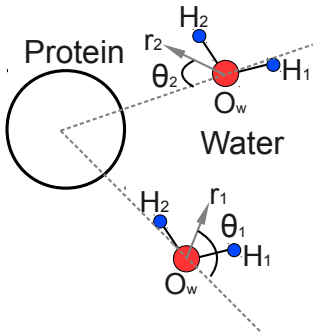

Fig. 7. Axes and orientation angles of water. The water atoms and unit vectors $\mathbf{r}_1$, $\mathbf{r}_2$ lie on the plane of the figure. $\mathbf{r}_3$ comes out of the page. Due to the rotational symmetry of a water molecule with respect to $\mathbf{r}_1$, θ_2 was taken to be the acute angle between $\mathbf{r}_2$ and the radially outward direction from the protein surface (dashed line). θ_3 was defined similarly.

Orientation angle. For the collagen system, the water orientation angle relative to the protein surface is the angle θ_i between $\mathbf{r}_i$ ($i = 1, 2, 3$) and the minimum distance vector from the nearest collagen helical axis to the water oxygen (Fig. 7). The helical axis of a peptide (triple helix) passes through its center of mass and is parallel to the x-axis in Fig. 5. For β-sheet bilayers, θ_i was measured between $\mathbf{r}_i$ ($i = 1, 2, 3$) and the direction perpendicular to the β-sheet (y-axis in Fig. 5).

By symmetry of the water molecule, θ_2 and θ_3 are in the range [0°, 90°]. A lower θ_2 and higher θ_3 correspond approximately to a radial orientation, as occurs around polar groups (Fig. 8d, Fig. 9d, Fig. 10g,h, Fig. 11g,h). For a randomly rotating water molecule, the average value for θ_1 ($0° \leq \theta_1 \leq 180°$) is 90°. On the other hand, the average of θ_2 or θ_3 are $\langle\theta_{2,3}\rangle = \frac{1}{2}[\int_0^{\pi/2} \theta \sin\theta d\theta + \int_{\pi/2}^{\pi} (\pi - \theta) \sin\theta d\theta] = 1$ radian. These correspond to the bulk values in the hydration maps.

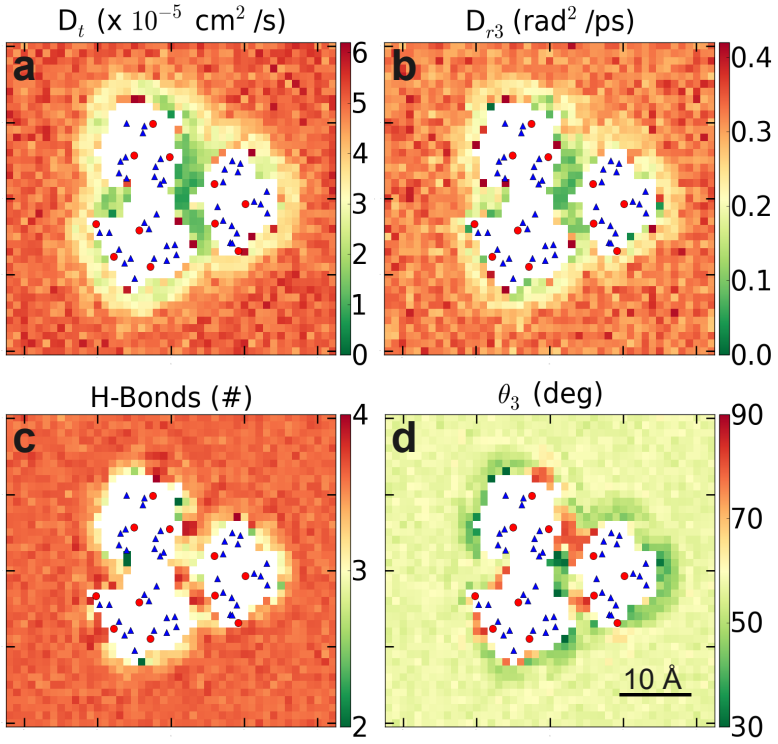

Fig. 8. A cross section of hydration maps for the three-peptide 2D3F system. Hydration maps for β-sheet filaments are in Fig. 10 and Fig. 11. Dots inside the protein (white region) are peptide oxygen (red circle) and carbon (blue triangle) atoms within ± 1.5 Å from the cross-sectional plane. (a) Translational (D_t) and (b) rotational (D_{r3}) diffusion coefficients. (c) Average number of hydrogen bonds. (d) Average radial orientation angle (θ_3) of water, which is overall larger near surface carbonyl oxygens. The bulk value of θ_3 approaches the analytical limit, 1 rad=57.3°. Fig. 9 shows maps for other quantities.

Number of hydrogen bonds. For frames where a water molecule visits the cell, the number of hydrogen bonds that a water molecule makes with the neighboring water molecules and the peptide atoms were counted with a hydrogen-oxygen distance cutoff of 2.4 Å[65], and averaged in each cell.

Hydration maps in the case without any constraint on the peptide was obtained by assigning a local coordinate frame to the cross section of the peptide [18] and defining unit cells with respect to the local frame (Fig. 12). The code for the hydration map analysis was custom written for the GROMACS simulation package (Appendix A).

4. Hydration Force

To measure intermolecular forces, heavy atoms on one side of the filament assemblies (Fig. 5, COL1 & COL2, βNP1, and βP1) were restrained with large spring constants $k = 10^4$ kJ/(mol·nm^2). The other part (COL3, βNP2, βP2) was translated along the y-axis by Δy_0 in 1-Å intervals. Δy_0 ranges [−4.0,+4.0] Å for collagen and [−2.0,+10.0] Å for β-sheets. Since collagens have a hydrated interface at the crystallographic separation ($\Delta y_0 = 0$ Å), a closer approach ($\Delta y_0 = -4.0$ Å) than the amyloid steric zipper interfaces ($\Delta y_0 = -2.0$ Å) was possible. At each Δy_0, C_α atoms of the translated part of the system were weakly constrained ($k = 500$ kJ/(mol·nm^2)). Denoting the number of these atoms by n_C and their average position during each 8-ns simulation by Δy, the net force per unit length of the molecule is $F_{tot} = n_C k(\Delta y - \Delta y_0)/(\text{peptide length})$ (dashed line in Fig. 13a-c). Hydration force, F_{hyd} (solid line), was obtained by subtracting from F_{tot} the average non-bonded interaction force F_{int} on the translated part (COL3, βNP2, and βP2) by the stationary part (Lennard-Jones and electrostatic; dotted line in Fig. 13a-c). Potential of mean force curves are shown in Fig. 14. To test the possible effect of the constraining potential on the measured force, for

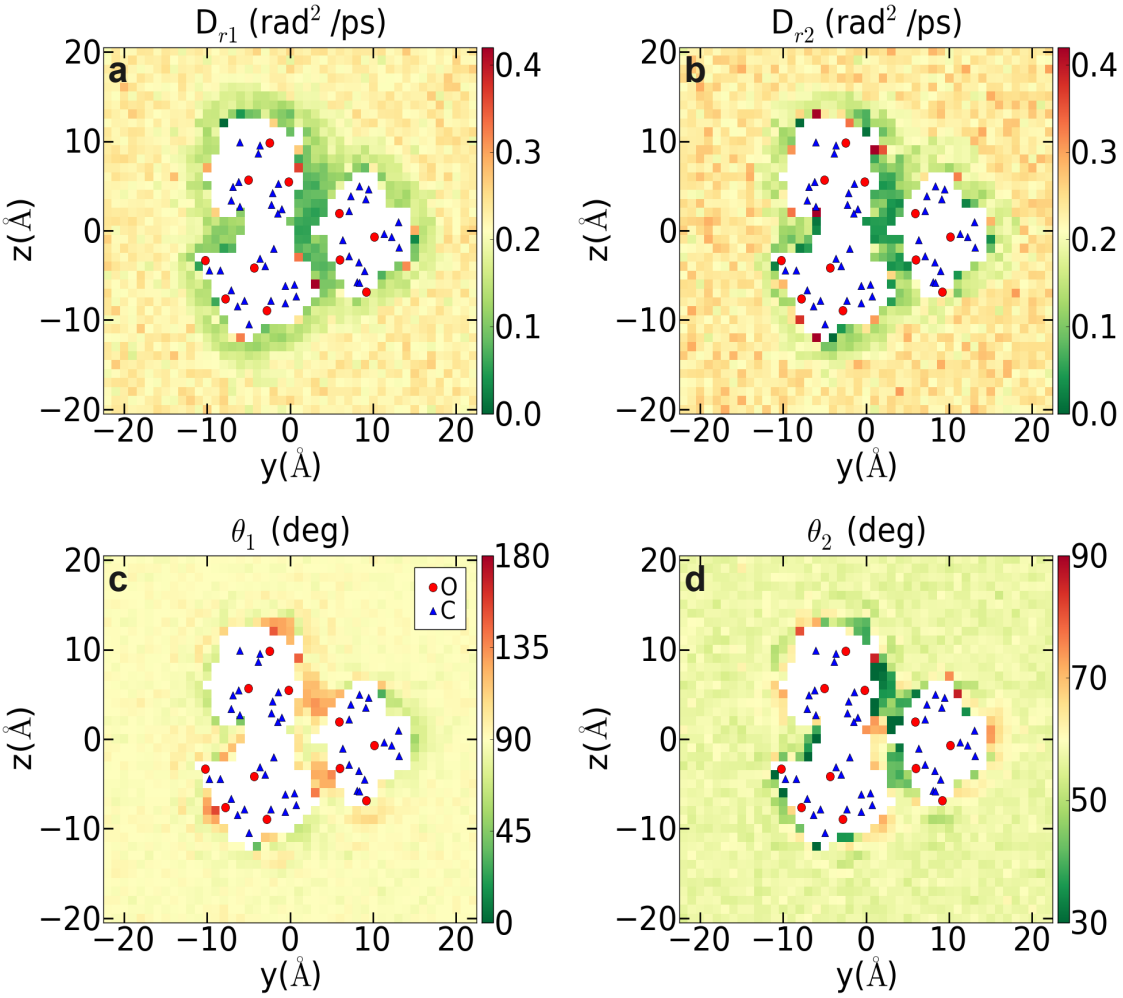

Fig. 9. Hydration maps for the rotational diffusion coefficients D_{r1}, D_{r2} and orientation angles θ_1 and θ_2 for PDB 2D3F. The cross-sectional plane is the same as in Fig. 8. While D_{r1} and D_{r2} are qualitatively similar to D_{r3}, the latter is higher (*cf.*, Fig. 8d), suggesting that the axis perpendicular to the plane of water atoms swivels the most. For a randomly rotating water molecule in bulk, the average value of θ_1 is 90°, and it is 1 radian ($\approx 57°$) for θ_2 and θ_3.

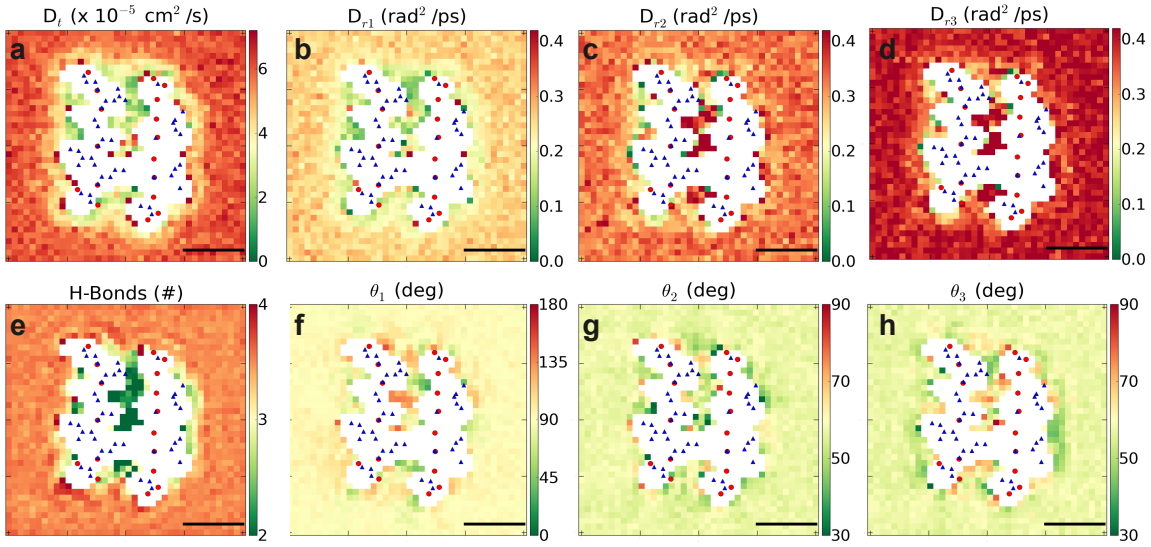

Fig. 10. Hydration maps for a β-sheet bilayer of PDB 2KIB. Despite the amino acid side chains being non-polar, the hydration maps are very similar to that of the hydrated collagen (Fig. 8 and Fig. 9). The translational and rotational diffusion coefficients are qualitatively similar to that of collagen, but are higher in magnitude since we used the TIP3P water model for β-sheet simulations (Fig. 6). The cross sectional plane is the same as in Fig. 4c. Scale bar is 10Å.

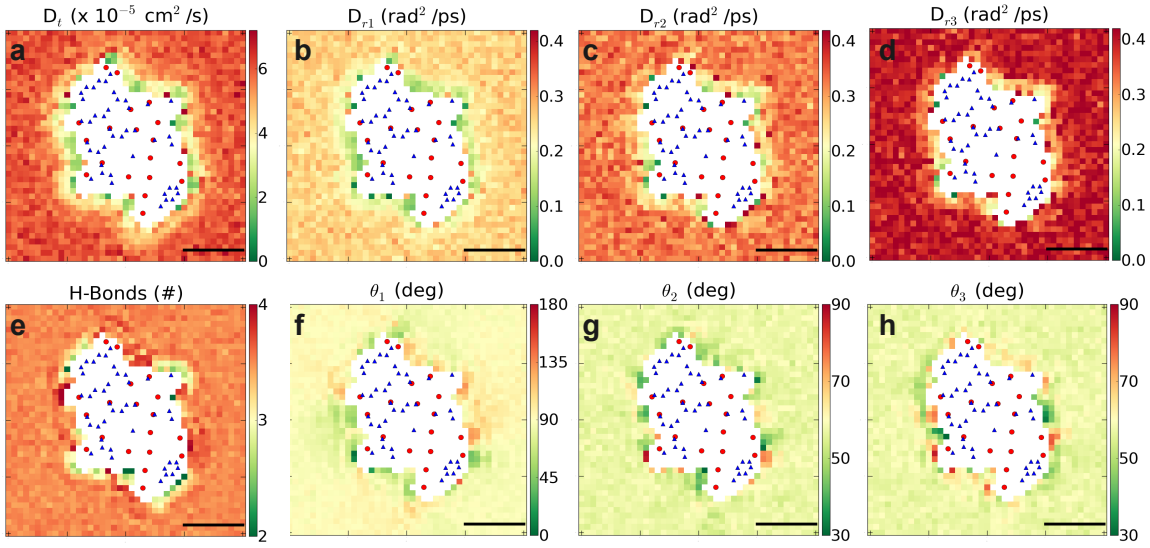

Fig. 11. Similar to Fig. 10, for PDB 1YJP, which has non-polar side chains. The cross sectional plane is the same as in Fig. 4d.

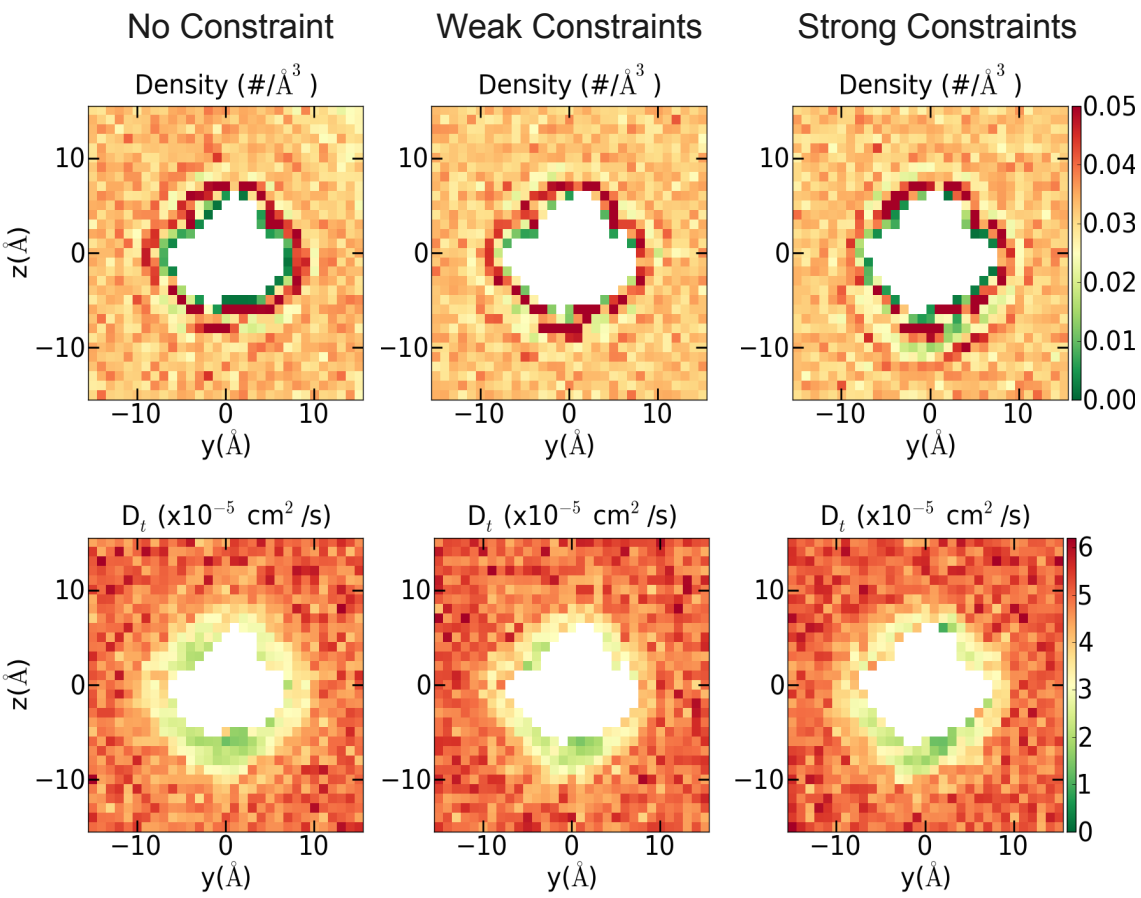

Fig. 12. Influence of protein motion on hydration maps. Hydration maps for a single 2D3F are shown at the same cross-section with varying harmonic constraints on protein C_α atoms, which are very similar. Spring constants for weak and strong constraints are respectively 10 and 10^4 kJ/(mol·nm^2).

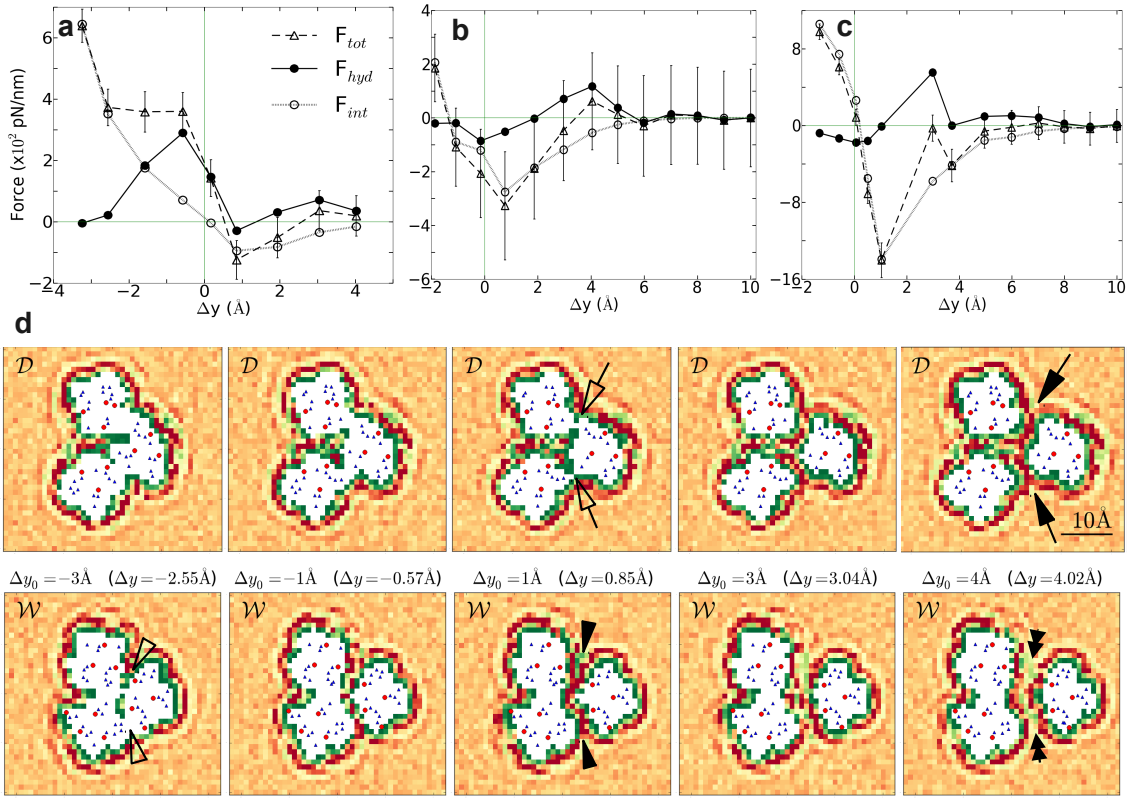

Fig. 13. Relation between hydration force and hydration map. (a-c) Forces per nanometer length of filament. (a) 2D3F (b) 2KIB, and (c) 1YJP. F_{tot}: net force, F_{int}: interaction force directly between peptides, and F_{hyd}: hydration force. Error bars (standard deviation) are shown for F_{tot}, which are comparable in magnitude to those for F_{int} and F_{hyd}. Although F_{hyd} oscillates, the corresponding energy profile does not (Fig. 14). (d) Water density maps shown in two cross sections at different separation distances of 2D3F. The color scale is the same as in Fig. 4. At the crystallographic separation ($\Delta y_0 = 0$ Å), $\mathcal{D}$ and $\mathcal{W}$ respectively denote regions with dry and wet interfaces between peptides (Fig. 20). Arrows and arrowheads indicate coalescence and depletion of primary hydration shells. Peaks of the oscillating hydration force in (a) are located between these transitions ($\Delta y_0 = -1$ and 3 Å). Similar maps for β-sheet filaments are in Fig. 19.

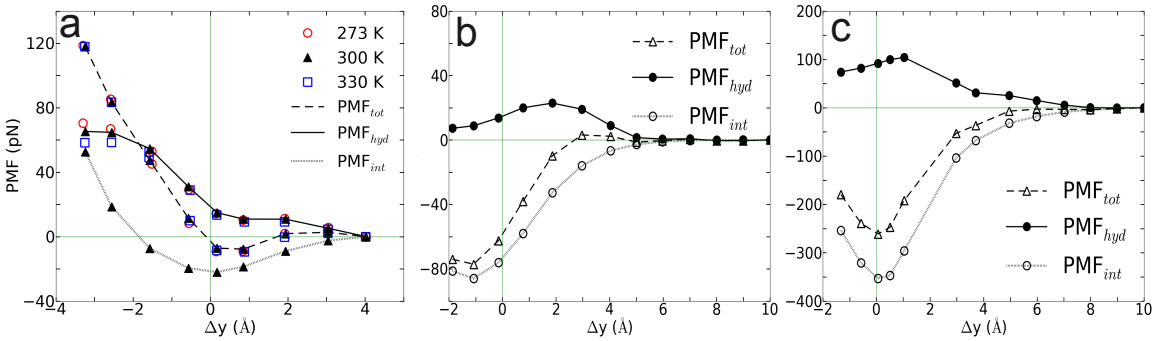

Fig. 14. Potential of Mean Force (PMF) per nanometer length of the filament, obtained by integrating the force curves in Fig. 13. (a) Collagen 2D3F; β-sheet peptides (b) 2KIB and (c) 1YJP (energy set to zero at $\Delta y_0 = 4$ Å in (a), and at 10 Å in (b), (c)). The PMF for F_{hyd} does not oscillate, contrary to previous suggestion for macroscopic surfaces [28].

2D3F, we varied their spring constants. The measured forces followed the same force profile, which suggests that the magnitude of constraints does not affect measuring the force that the molecule experiences (Fig. 15).

When calculating the average separation Δy for a 8-ns simulation, we excluded the initial period during which the translated part (Fig. 5) reaches its equilibrium position. For 93% of simulations, this time was less than 50 ps, except for some cases of 2KIB and 1YJP, where it took 0.2–4 ns to reach the equilibrium position. Force profiles obtained from the first and the last half of the measured intervals were nearly identical, indicating sufficient sampling time. Likewise, hydration maps were calculated based on the trajectory after the initial transient.

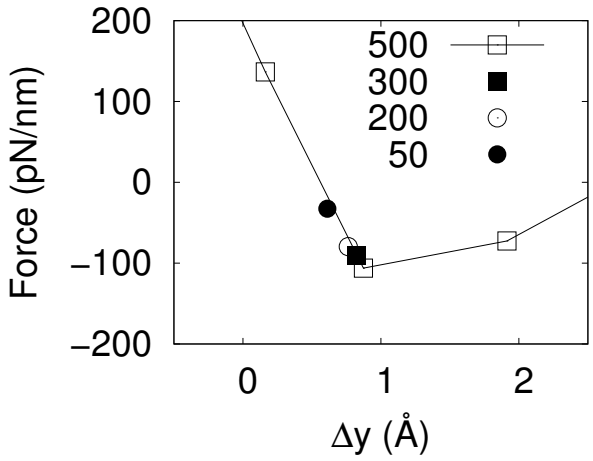

Fig. 15. Dependence of the measured force on constraint used in simulation. Legends are spring constants in kJ/(mol·nm^2) used for the constraint on C_α atoms of COL3 in Fig. 5a at 330 K. Using a weaker constraint results in a greater displacement from the initial position Δy_0 (= 1 Å), *i.e.*, a smaller average position Δy. The calculated forces lie on the same curve, suggesting that the strength of the harmonic constraint does not affect our measurement.

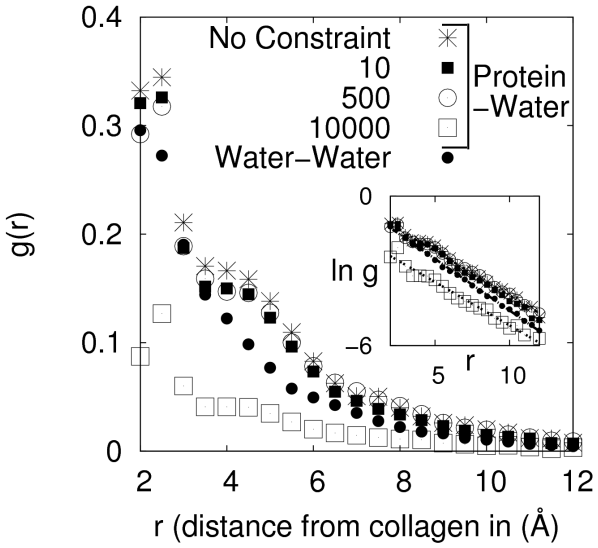

Fig. 16. Protein-water and water-water correlation functions in thermal motion measured for a single PDB 2D3F. We use Eq. 1 in Methods to calculate $g(r)$. Numbers in legend are spring constants k of the constraints on C_α atoms in kJ/(mol·nm^2). $g(r)$ decreases when $k = 10^4$ kJ/(mol·nm^2). However, with weaker or zero constraints, $g(r)$ is nearly identical to that for the water-water correlation. Thus, water motion near the protein surface is only as correlated as with thermal motion of other water molecules in the bulk. Inset: Log-normal plots of $g(r)$ revealing an exponential decay. The characteristic decay lengths are 2.72±0.056 Å (no constraint), 2.60±0.052 Å (k =10 kJ/(mol·nm^2)), 2.80±0.055 Å (500 kJ/(mol·nm^2)), 2.87±0.10 Å (10^4 kJ/(mol·nm^2)), and 2.42±0.027 Å for water-water correlation. They are all similar to the thickness of a hydration shell, suggesting that the already weak correlation in thermal motion becomes negligible after the primary hydration shell.

5. Local Correlation Function in Thermal Motion

Using PDB 2D3F, we tested the correlation between protein motion and the motion of nearby water molecules. We followed displacements of a water oxygen and the nearest collagen heavy atom during 1 ps and assigned unit vectors in the corresponding directions, denoted as $\mathbf{u}$ and $\mathbf{a}$, respectively. With r representing the distance between the two atoms, we define the correlation function

$$g(r) = \langle \mathbf{u} \cdot \mathbf{a} \rangle_r \tag{2.1}$$

where the average is over all water oxygen–nearest protein heavy atom pairs within the range $(r, r + \Delta r)$ ($\Delta r = 0.5$ Å). For comparison, we randomly selected 20 water molecules in bulk and measured $g(r)$ for each with the surrounding water molecules, and averaged them to obtain the water–water correlation function (Fig. 16).

6. Effect of Water Models on Hydration Maps

We compared hydration maps of the smaller peptide 1A3I using three different water models, SPC, TIP3P, and SPC/E. The density maps are nearly identical (Fig. 6). The translational diffusion coefficient (D_t) of the SPC/E water in bulk is the closest to the experimental value ($\approx 3.1 \times 10^{-5}$ cm^2/s) while it is higher for SPC ($\approx 4.8 \times 10^{-5}$ cm^2/s) and TIP3P ($\approx 5.8 \times 10^{-5}$ cm^2/s), as reported earlier [64]. Similar differences in magnitude were also seen for rotational diffusion coefficients. But their profiles near the protein surface are qualitatively similar (Fig. 6).

C. Results and Discussion

1. General Features of the Hydration Shell

To distinguish water-mediated interactions from those caused by conformational motion, we harmonically constrained the peptides, and calculated local water density, diffusion coefficients (translational and rotational), orientation angles relative to the protein surface, and the number of hydrogen bonds in 1-Å resolution (see Methods). The resulting *hydration maps* (Fig. 17, Fig. 4, and Fig. 8) were nearly identical with or without the harmonic constraint (Fig. 12). This is likely because hydration water organizes faster than the conformational motion of the filaments. Furthermore, directions of thermal motion of water oxygen and nearby protein heavy atoms were only weakly correlated (Fig. 16). Constraining or fixing protein atoms have been previously used for elucidating the behavior of the hydration water [49, 45]. The results below are thus obtained with restraints on proteins, which also allows more extensive sampling and intermolecular force measurement.

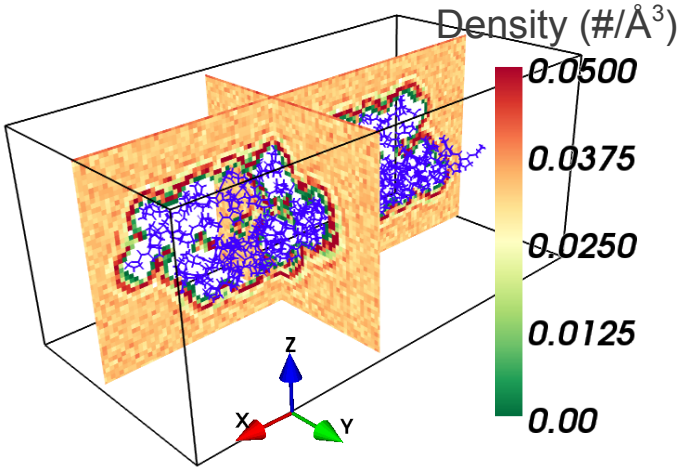

Fig. 17. Simulation setup and density hydration maps of a three-peptide collagen system. The peptides (Protein Data Bank (PDB) ID 2D3F [24]) have a hexagonal packing in the crystal, similar to native fibrillar collagens [66]. Two cross sectional density hydration maps are shown. A well-defined primary hydration shell (red) around the peptides is visible, followed by a less dense secondary hydration shell (*cf.*, Fig. 4a). The coordinate frame defined here is used in all other figures.

In all the systems tested, the primary hydration shell formed regardless of the type of the underlying amino acid (Fig. 4). Instead of forming a depletion zone [28], high-density hydration shell also formed around the non-polar PDB 2KIB, as has been observed for a model spherical hydrophobe [51], also known as the hard-wall effect [55, 56]. Translational and rotational diffusion coefficients are 2–5 fold lower across the protein surface than in bulk water, and increase monotonically away from the surface (Fig. 8a,b, Fig. 9a,b, Fig. 10a-d, and Fig. 11a-d), which is consistent with previous computational [12, 46, 67, 68] and experimental studies [69, 70]. These results suggest

that the formation of primary hydration shells and retardation of water motion are mainly due to boundary-induced packing effects rather than by specific interactions between water molecules and particular surface groups. In the case of PDB 2KIB, which has an anti-parallel β-sheet structure, there is a low density of water within the bilayer (Fig. 4c) indicating that the two β-sheets do not form the steric zipper as tightly as in PDB 1YJP that is a parallel β-sheet (Fig. 5) [71]. When we performed simulation without any constraint on the filaments, the two β-sheets of 2KIB packed more closely, eliminating the interfacial water. The translational diffusion coefficient for the inter-sheet water molecules is low (Fig. 10a), while their rotational diffusion coefficients are nearly as high as in the bulk (Fig. 10b-d). Though we have lesser statistics in the low density regions, the high rotational diffusion is likely a result of the absence of hydrogen bonds within the non-polar interface (Fig. 10e).

Unlike the density and diffusion maps, the average number of hydrogen bonds between water and the protein surface is non-uniform, which is higher around polar groups and lower around non-polar groups (Fig. 8c, Fig. 10e, Fig. 11e). Retardation of water motion near the protein surface even with a smaller average number of hydrogen bonds suggests a glassy rather than ice-like state [13]. As a result of hydrogen bond formation, water orients with one of its O–H bonds aligned radially outward from the surface around polar groups, while the orientation is circumferential near the non-polar groups, which can be seen near the surface oxygen atoms (red circles) in Fig. 8c and Fig. 9c,d [12, 23, 22]. For β-sheet filaments, the trend is similar, but less clear (Fig. 10f–h and Fig. 11f–h). This is because the surface normal direction defined relative to the plane of the β-sheet (see Methods) has poorer correlation with water orientation due to the bulkier side chains, whereas the collagen triple helix of PDB 2D3F has a smoother surface geometry with less bulky prolines and hydroxyprolines.

2. Hydration Force Profile

We measured the intermolecular force on one collagen triple helix or a β-sheet layer translated perpendicular to the filament axis (Fig. 5; see Methods). The net force F_{tot} (in the y-direction) consists of the interaction force F_{int} directly between proteins, and the water-mediated (hydration) force F_{hyd} (Fig. 13). The profile of F_{hyd} shows an oscillation whose amplitude decays with Δy. The 3$\sim$4 Å oscillation period of F_{hyd} is comparable to the diameter of a water molecule, similar to the behavior between larger surfaces [28]. It should be noted that oscillation is present even between non-polar surfaces (Fig. 13b). But the location and magnitude of the hydration barrier (maximum of F_{hyd}) depends on the type of the surface. In the case of collagen that remains hydrated after assembly, the barrier occurs below the crystallographic separation (Fig. 13a; $\Delta y = -0.57$ Å). For β-sheet bilayers that form dry interfaces, hydration barriers are located further away (Fig. 13b,c; $\Delta y = 3.7 \sim 4.0$ Å). The polar 1YJP has the greatest hydration barrier (555 pN/nm; force is measured per nm length of the filament), followed by 2D3F (289 pN/nm). Nevertheless, 1YJP forms a dry interface since the barrier is formed further away than the location of the maximum attraction of the dominant force, F_{int} (van der Waals and electrostatic; $\Delta y = 1.14$ Å).

Between the amyloid filaments, $F_{hyd} < 0$ near $\Delta y \sim 0$ Å, indicating that the surrounding water tends to prevent dissociation of β-sheets as the dry steric zipper interface is formed. This occurs for both 1YJP and 2KIB that respectively have polar and non-polar interfaces. Thus, ironically, the 'hydrophobic' force between the β-sheets in close separation may be regarded as a manifestation of the hydration attraction. Also, due to the flat geometry of β-sheet bilayers, the attractive force per nanometer length of the 2KIB filament ($F_{hyd} = -89$ pN/nm at $\Delta y = -1.08$ Å) is

greater than that for the cylindrical and hydrated collagen ($F_{hyd} = -30$ pN/nm at $\Delta y = 0.85$ Å), and it is the greatest for 1YJP ($F_{hyd} = -165$ pN/nm at $\Delta y = 0.06$ Å) whose side chains form a better steric zipper interface than 2KIB (Fig. 13a–c). For PDB 2KIB, the maximum hydration attraction occurs at $\Delta y = -1.08$ Å, which is consistent with the expulsion of water between the β-sheets in simulations without any constraint.

The dominance of F_{int} over F_{hyd} in β-sheet bilayers agrees with an earlier finding that interaction energy and geometric complementary, rather than solvation free energy, are major contributors for stabilizing the amyloid steric zipper structures [71]. By contrast, in the case of the collagen peptide PDB 2D3F, since the major hydration barrier is located at the rising phase of the F_{int} (Fig. 13a), the interface remains hydrated after assembly. These results suggest that the three types of surfaces differ in the amplitude of the oscillating and decaying hydration force, as well as in the *phase* of oscillation relative to the equilibrium distance.

For force measurements we used positional restraints to maintain the filaments in the original straight conformations as in the PDB files (Fig. 5). The straight conformations are due to the packing effect of filaments [72], which may not be the lowest in energy for small systems as in Fig. 5. Without the restraint, they develop super-helical twist. For collagen, the three peptides of 2D3F in Fig. 5a mildly twist together, whereas the β-sheet bilayers become self-twisted, as reported previously [73, 26]. However, the deformation makes it difficult to define the inter-filament distance. Since our measurement reports the behavior of water for a given conformation of the filament, whether or not the filament is in the minimum energy conformation does not affect our main conclusions. As a result, the net force (F_{tot}) is non-zero at the original separation in the PDB file, $\Delta y_0 = 0$ Å. For 2D3F, forces in the transverse directions (x- and z-directions) are much smaller than in the longitudinal direction (y-direction)

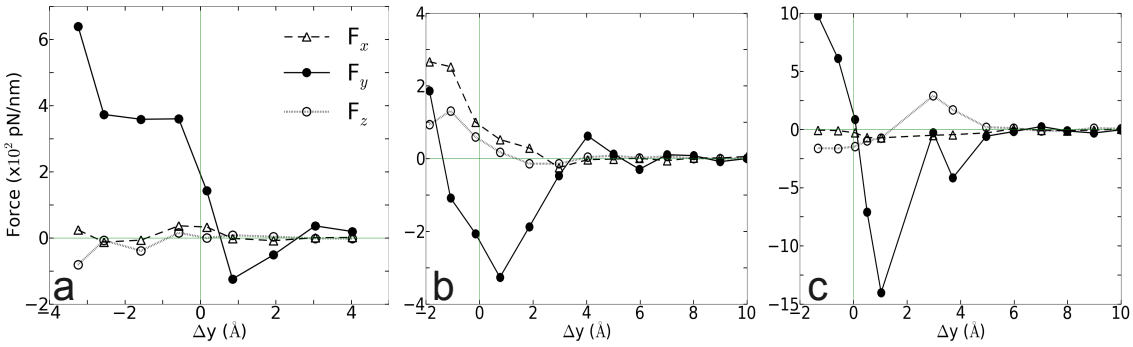

Fig. 18. Comparison of the net forces in three directions for (a) 2D3F, (b) 2KIB, and (c) 1YJP. Coordinate system is defined in Fig. 5a. Compared to F_y (F_{tot} in Fig. 13a-c), forces in the orthogonal directions (F_x and F_z) are much smaller (except in (b) 2KIB when $\Delta y < 0$ Å). Higher F_z at $\Delta y \approx 4$ Å in (c) 1YJP is due to interaction between the longer Q4 side chains.

(Fig. 18a). This is also the case for PDB 2KIB, except for $\Delta y_0 \leq 0$ Å owing to the poor surface complementarity (Fig. 5b; Fig. 18b). In 1YJP, the transverse forces are much smaller than the force in the longitudinal direction except for $\Delta y = 3 \sim 4$ Å (Fig. 18c), which is due to the attraction between the Q4 side chains in the z-direction (*cf.*, Fig. 19b, white region between the two β-sheets at $\Delta y_0 = 4$ Å).

3. Structural Origin of the Hydration Force

Comparing the force profiles and the corresponding hydration maps reveals that the oscillation of F_{hyd} correlates strongly with the coalescence and depletion of hydration shells. In the case of collagen, there are both dry ($\mathcal{D}$) and wet ($\mathcal{W}$) interfaces at $\Delta y_0 = 0$ Å (Fig. 20). The region forming $\mathcal{D}$ has a single hydration shell shared between surfaces at $\Delta y_0 = 4$ Å (solid arrow in Fig. 13d), which disappears by $\Delta y_0 = 1$ Å (open arrow). Similarly, in the region forming $\mathcal{W}$, two separate hydration

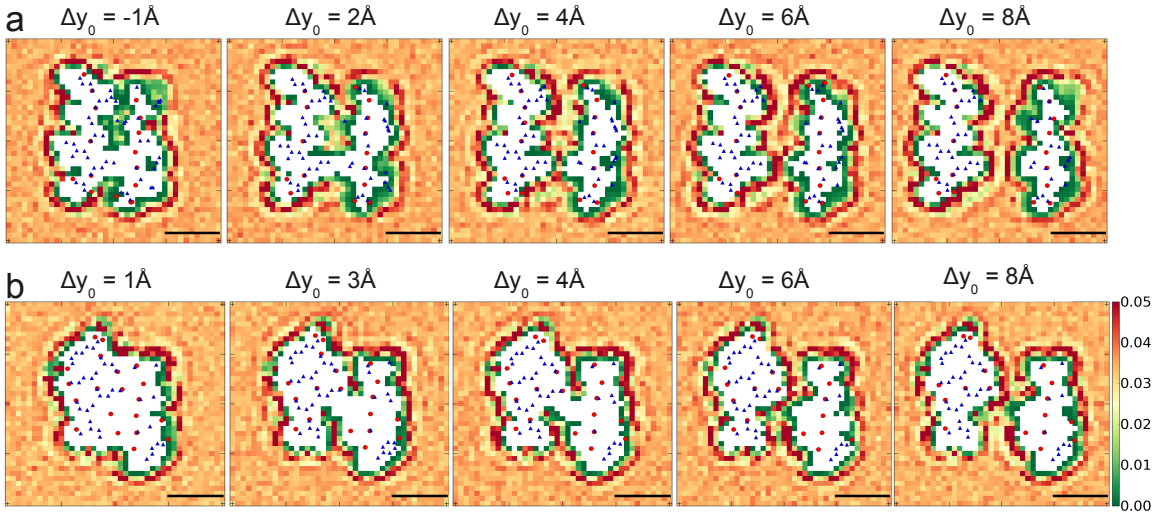

Fig. 19. Distance-dependent changes in hydration shells between β-sheets (*cf.*, Fig. 13d). (a) 2KIB (b) 1YJP. Since only the backbone C_α atoms are constrained in β-sheets on the right side (βNP2, βP2 in Fig. 5b,c), movement of side chains leads to blurring of the density hydration map and hence the low-density region (green). Because of this, coalescence and depletion of hydration shells are not as clear as in the case of collagen that has less bulky side chains (Fig. 13d).

shells ($\Delta y_0 = 4$ Å, double arrowhead) merge into one ($\Delta y_0 = 1$ Å, solid arrowhead), which eventually disappears ($\Delta y_0 = -3$ Å, open arrowhead). Hydration shells disrupt between these transitions at which the hydration repulsion is maximal ($\Delta y_0 = -1$ and 3 Å). But when the barrier is overcome, bias towards completing the transition causes an attraction ($F_{hyd} < 0$).

The correlation between the oscillation of F_{hyd} and the coalescence/depletion of hydration shells are also present in β-sheet bilayers, although it is not as clearly seen due to the motion of the bulkier amino acid side chains (Fig. 19). In the case of 2KIB,

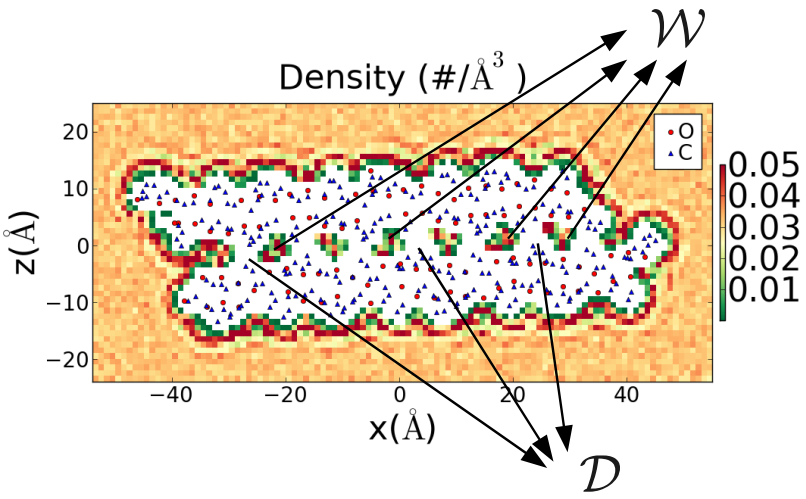

Fig. 20. Alternating dry ($\mathcal{D}$) and wet ($\mathcal{W}$) regions between COL1 and COL2 in Fig. 5a. Similar interfaces are formed between COL1 and COL3, and COL2 and COL3 at the crystallographic separation.

a low density depletion zone is formed between the β-sheets at close separations (Fig. 19a, $\Delta y_0 = 2$ Å), below which hydration force becomes attractive (Fig. 13b). This is consistent with previous reports on the dewetting transition in confined non-polar surfaces [47, 33]. However, due to the robust formation of hydration shells, dewetting does not occur at greater separations.

In contrast to the distance-dependent behavior of the hydration shells, maps for the number of hydrogen bonds or water orientation did not show any clear correlation with the hydration force profile, which we checked by going over multiple cross sections of the filament in addition to those displayed in Fig. 13 and Fig. 19. Water diffusion coefficients near the protein surfaces are low in all cases. Thus, coalescence and depletion of primary hydration shells are the major determinants for the oscillatory profile of hydration force, which occurs regardless of the type of the surface. While it was necessary to apply harmonic constraints to the peptides in order to calculate forces as functions of the separation distance, since the force profile does not depend on strengths of the constraints (Methods), in their absence, major features of F_{hyd} would be preserved locally. However, conformational motion of the molecule without a constraint will make it difficult to single out the effect of the hydration force, and it may even appear to be monotonic when averaged over the length of the filaments that have non-uniform surface separations.

Whereas the oscillation in F_{hyd} is due to the interaction between the primary hydration shells, multiple factors may affect its amplitude and phase, such as surface geometry and local hydrogen bonding events. To further illustrate the nature of the hydration force, we calculated the force profiles at 273, 300, and 330 K (Fig. 21 and Fig. 22). The hydrated PDB 2D3F has the least temperature dependence, followed by the polar 1YJP and then the non-polar 2KIB (Fig. 21). As seen in Fig. 22b,d,f, variations in F_{int} at different temperatures is higher at shorter distances due to the

stronger interaction between side chains. For β-sheets, this can be seen by the root-mean-square fluctuation of the unconstrained side chain atoms facing the interface in the range $\Delta y_0 \leq 0$ Å. This is 0.86–1.22 Å (273 – 330 K) for 2KIB and 0.33–0.39 Å for 1YJP, which corroborates with the size of temperature-dependent variations in F_{int}. However, at larger separations, F_{int} is independent of temperature, even though the side chain motion increases further. Even for 2KIB that has the greatest side chain motion (possibly because side chains in a peptide do not flank others in the neighboring ones with the anti-parallel β-sheet arrangement), F_{int} is insensitive to temperature beyond $\Delta y_0 \geq 3$ Å (Fig. 22d). Thus, temperature dependence of the hydration force calculated in this range is more reliable than at shorter distances. For 2KIB, the hydration barrier appears to decrease with temperature, suggesting that its hydration shell becomes easier to disrupt with increasing temperature. While this may be due to the lack of hydrogen bonds between water and the non-polar surface of PDB 2KIB, since the three systems we test are not identical in surface topography, geometry-related effect on the stability of the hydration shell cannot be ruled out.

Since the CHARMM force field we used [58] is non-polarizable, to test whether forces are affected by temperature-dependent changes in water dipole moment [74], for PDB 2D3F, we decreased partial charges of the water model by 5%, which yielded no major difference (cross in Fig. 21a and Fig. 22a). Despite an ∼2-fold increase (by 95%) in diffusion coefficients of water molecules in bulk from 273 K to 330 K the measured hydration forces show little temperature dependence. Furthermore, the profile of F_{hyd} remained nearly the same when the SPC/E water model was used (diamond in Fig. 21 and Fig. 22). Since the SPC/E water has lower diffusion coefficients than the other water models we used (Fig. 6), this result further supports that hydration force does not depend on translational or rotational motion of water molecules. Although temperature dependence may become more pronounced for larger systems, possibly

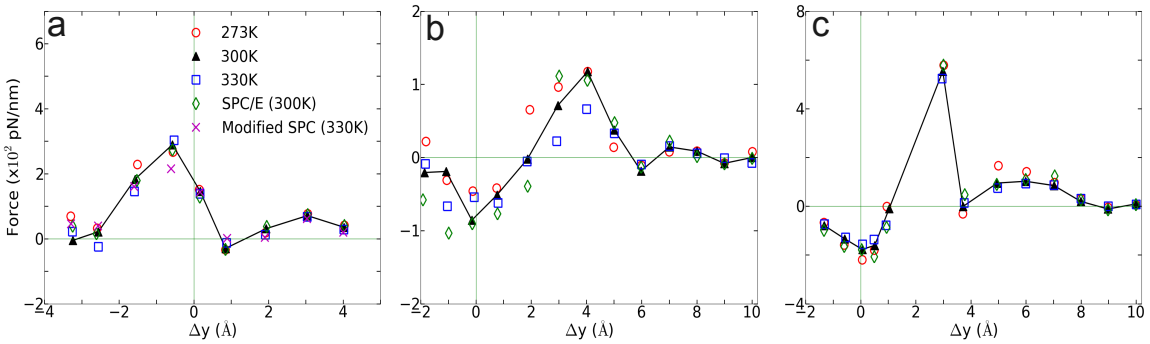

Fig. 21. Comparison of hydration forces measured at different temperatures or with different water models. (a) 2D3F, (b) 2KIB, and (c) 1YJP. Temperature dependence (273, 300, and 330 K in legend) was with (a) SPC and (b,c) TIP3P water models. For comparison, forces were also measured using the SPC/E water at 300 K (diamond). In (a), the water oxygen and hydrogen atoms of the modified SPC (cross) has 95% of the partial charges of those in the regular SPC water, to mimic the reduction in water dipole moment at high temperature [74]. The corresponding profiles of F_{tot} and F_{int} are in Fig. 22.

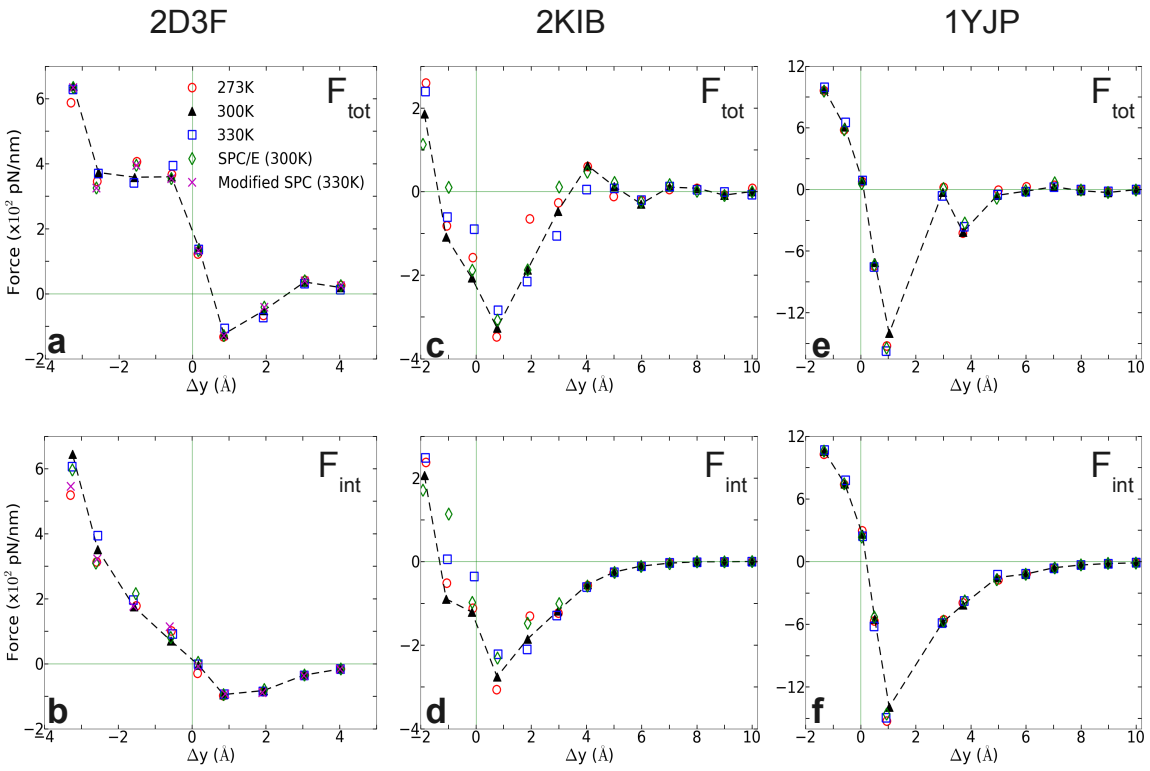

Fig. 22. Force curves (F_{tot} and F_{int}) at different temperatures or with different water models. (a,b) 2D3F, (c,d) 2KIB, and (e,f) 1YJP. Legends are the same as in Fig. 21.

for PDB 2KIB, the present results suggest that the main determinant of hydration force is the interaction between hydration shells formed by surface-induced packing of water molecules. The experimentally measured temperature dependence of the force between collagens, DNAs, and polysaccharides [53, 75] may instead be due to conformational motion where closer parts of the molecules interact more strongly while the average intermolecular distance is farther away, which would make the attraction appear to increase with temperature [53, 75].

D. Conclusion

The present results show that the previously postulated role of the hydration shell in generating hydration forces [20, 38] is to some extent applicable to the self-assembly of filamentous proteins studied here. However, we find that the hydration shell is formed ubiquitously over all surface types in the systems tested, where difference lies in the magnitude and the location of hydration barrier and hydration attraction, which can also be affected by the surface geometry or complementarity [49, 28]. The similarity between the behaviors of the hydration water near hydrophobic and hydrophilic protein surfaces has been previously suggested [68], although to our knowledge, its implication in hydration force has not been addressed in detail. Thus, designating a protein surface as either 'hydrophobic' or 'hydrophilic' may be too simplistic of a dichotomy, while surfaces in reality lie between these idealized limits [76]. There should be no fundamental difference in the way hydration forces arise among different types of protein surfaces even with varying affinity with water. [41]. Whether two surfaces eventually bind or repel will be determined by the relative magnitude and phase of hydration and interaction forces, as well as surface complementarity and flexibility. While the behavior of the hydration water may be more complex for small flexible

peptides or globular proteins that have mobile sub-domains [77, 78, 79], the present results are likely applicable to a broad range of protein complexes or assemblies whose interfaces are geometrically similar to those studied here.

CHAPTER III

UNDERSTANDING OPTICAL CLEARING OF COLLAGENOUS TISSUES THROUGH MOLECULAR DYNAMICS SIMULATIONS*

A. Introduction

Biocompatible chemical agents have been shown to induce a temporary and reversible reduction in tissue light scattering [80, 81]. With concentrated sugar-alcohols such as sorbitol and sugars such as high fructose corn syrup (HFCS), light scattering can be reduced by as much as five-fold. These chemical agents have been observed to be most effective in vitro when applied directly to the mesenchyme, e.g., the dermis of skin, and less so when applied topically. The putative mechanism of optical clearing is index matching of tissue light scatterers via optical immersion. Index matching as a mechanism is intriguing because, ultimately, a reduction in light scattering must coincide with a homogenization of the index of refraction. Yet, as a physical parameter, index of refraction of potential optical clearing agents cannot be used to predict 'clearing' ectiveness. Tissue dehydration has also been proposed as a mechanism of optical clearing. Water makes up a substantial proportion of tissue weight and has an index of refraction significantly different from that of insoluble biological constituents. However, chemical agent osmolarity cannot be used as a predictor of its optical clearing potential. A better understanding of how the index of refraction

*Portions of this chapter are printed with permission from J. M. Hirshburg, K. M. Ravikumar, W. Hwang, and A. T. Yeh, "Molecular basis for optical clearing of collagenous tissues." *J Biomed Opt*, vol. 15, no. 5, p. 055002, 2010. Copyright 2010 by Society of Photo-Optical Instrumentation Engineers (SPIE).

dx.doi.org/10.1117/1.3484748

is homogenized in optically cleared tissue could lead to a rational basis for designing effective, clinically applicable formulations.

Previously, it has been shown at microscopic and ultrastructural length scales that glycerol, a prototypical optical clearing agent, destabilizes high-order collagen structures and that this effect coincides with agent-induced tissue transparency [82, 83]. It has been suggested that collagen destabilization was due to the chemical agent's ability to screen non-covalent attractive forces. These same forces drive collagen fibrillogenesis from solution and have been characterized in the presence of sugars and sugar-alcohols [84]. Hirshburg and Yeh introduced collagen solubility as a measure of a chemical agent's ability to screen non-covalent forces and correlated it with tissue optical clearing for a series of polyols and sugars [85, 84, 86]. These studies suggested that, of the chemical agents in the series, their optical clearing potential (and collagen solubility) could be grouped by chain length and locations of hydroxyl groups within the molecule. More specifically for sugar-alcohols, optical clearing potential for the series could be grouped by chain length, in increasing order, as two-carbon backbone (ethylene glycol), three- to five-carbon backbone (glycerol to xylitol), and six-carbon backbone (sorbitol). We examine the interactions between clearing agents and collagen using a combination of MD simulations and experiments (experiments were done in Dr. Yeh's lab, Texas A&M University, by Jason Hirshburg and Alvin Yeh [87]) to elucidate the roles of collagen solubility in skin optical clearing. Our results suggest that the position of hydroxyl groups on alcohols impose steric constraints for forming surface bridges on the collagen triple helix. Thus it is an important factor in determining its optical clearing properties. Combined with experimental results, our simulation provides an atomistic picture for the non-linear trend in optical clearing properties of alcohols as their chain length increases.

B. Molecular Dynamics Simulations

For simulation, we used CHARMM version 34 with param22 force field [88]. The Generalized Born with a Simple sWitching (GBSW) implicit solvent model was used to account for solvation effects [89]. The main focus of our simulation was finding geometric constraints and propensities of various polyols in forming hydrogen bond bridges on collagen surfaces that are determined mainly by the structures of these molecules. Thus using an implicit solvent model rather than more computationally demanding explicit water simulation was sufficient for our purpose.

We used synthetic peptides 1BKV (Protein Data Bank ID) and a regular GPO peptide, $((GPO)_{10})_3$) (G-Glycine, P-Proline, and O-Hydroxyproline) in our simulations. Peptide 1BKV has a biologically relevant imino acid deficient region of type III collagen, which is important in collagen cleavage [18, 11]. Peptide GPO forms the most stable triple helical motif [90, 11]. Its backbone structure was built using the TheBuSr collagen building script [91]. Side chain atoms were added to the backbone using the existing amino acid topology files and systematically energy minimized as done previously [11] to get the final structure. Polar hydrogens were added to the peptides using the HBUILD facility in CHARMM [92]. Parameters for hydroxyproline were added from a previous study [59]. Structure of collagen triple helix constructed using this method agreed very well with existing x-ray data [18]. Structures of the sugar-alcohols (glycerol, xylitol, and sorbitol) were built from the already existing lipid topology and parameter files.

Each alcohol was simulated with either 1BKV or the GPO triple helix, making a total of 6 separate simulation runs. In each simulation, alcohol molecules were placed randomly around the peptide at a radial distance of 12.0 Å from the cylindrical axis of the triple helix. Owing to their bigger molecular structures, only 12 molecules of

xylitol and sorbitol were placed around the peptide, while 20 glycerol molecules were used. The systems were first energy minimized to remove close contacts and then heated at the rate of 5 K/ps for 60 ps to 300 K. The systems were then equilibrated for 40 ps at 300 K. Each production run was performed for 600 ps using the Verlet integration algorithm with a time step of 2.0 fs. As the system was in a microcanonical ensemble (constant total energy), binding of alcohols to collagen resulted in a slight increase of temperature due to the loss of potential energy. The maximum variation was seen in the case of sorbitol simulation where the average temperature was 309Å$\pm$6 K. Note that this does not affect our results, since we only study the hydrogen bonding modes of different alcohols which will not change appreciably within the range of temperatures.

Coordinates were saved every 1 ps. To prevent the drift of the peptide outside the simulation boundary (see below), harmonic constraints (spring constant = 2 kcal/(mol Å^2) were applied on all peptide atoms to their original positions during heating and equilibration. During the production run, harmonic constraints were applied only on backbone amide nitrogen, alpha carbon, and carboxylic carbon atoms, leaving the rest of the atoms, including the side chains, to freely interact with the alcohol molecules. To prevent diffusion of alcohol molecules away from the peptide, a cylindrical boundary 40 Å in diameter was imposed around the peptide, which activated a radial harmonic potential when alcohol molecules moved beyond the boundary, with a spring constant of 1 kcal/(mol Å^2).

Hydrogen bonds were defined using a distance cutoff of 2.4 Å and a hydrogen bond bridge, which we will simply refer as a bridge, was defined to be formed if an alcohol molecule was simultaneously hydrogen bonded to two collagen atoms (Fig. 23). We analyzed the propensities of the hydroxyl groups of alcohols to form bridges with collagen peptides as follows. Bridges formed by alcohol -OH groups on carbon

positions on adjacent positions were labeled type I; bridges formed between positions separated by one carbon were labeled type II and so on. Bond multiplicity (n) of each type was also considered. For example in the three-carbon glycerol molecule, a type I bridge with collagen can involve hydroxyl groups in positions {1,2} or {2,3} leading to an n value of 2, while type II bridge (Fig. 24) involves the hydroxyl groups in positions {1,3} with an n value of 1.

C. Results

To understand the hydrogen bonding interactions of alcohols on the collagen surface, we performed MD simulations of collagen peptides 1BKV (containing imino-poor domain) and GPO (imino-rich) with glycerol, xylitol, and sorbitol. Hydroxyl groups of alcohols can form hydrogen bonds with collagen atoms and can displace water molecules in the hydration shell. It is expected that the disruption of the hydration shell is greater if the alcohol bridges span more extensively across the collagen surface (Fig. 23). Thus, propensity to form a bridge will be an important factor in optical clearing.

Bridging propensities of hydroxyl groups in glycerol, xylitol and sorbitol are shown in Fig. 24. Bridges were categorized by the positions of the participating hydroxyl groups. Type I bridges were formed with hydroxyl groups on adjacent carbon atoms (hydroxyl positions {1,2},{2,3},{3,4}, etc), type II bridges were formed with hydroxyl groups separated by one carbon atom (hydroxyl positions ({1,3},{2,4} etc), type III bridges were formed with hydroxyl groups separated by two carbon atoms ({1,4},{2,5} etc), and so on. Note that higher bridge type numbers can screen collagen-collagen and collagen-water interactions more effectively than lower bridge type numbers (Fig. 23).

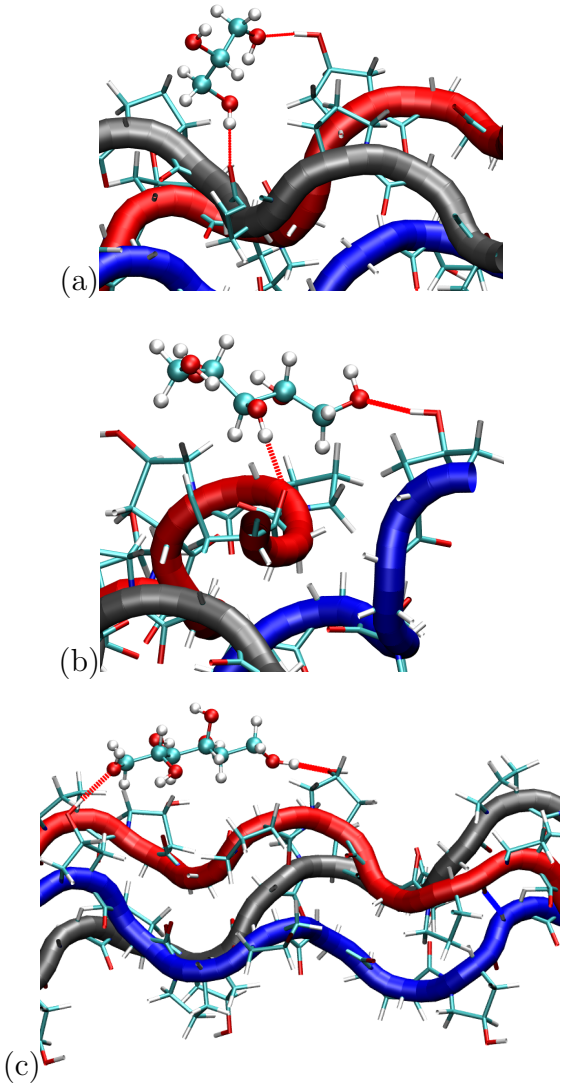

Fig. 23. Typical hydrogen bond bridges in alcohols. Bridge of -OH groups between (a) one and three carbon positions (type II) in glycerol, (b) one and three carbon positions (type II) in xylitol, and (c) one and five carbon positions (type IV) in sorbitol. Higher bridge types, as in (c), span further across the collagen surface and can potentially disrupt collagen-collagen and collagen-water interactions better than lower bridge types.

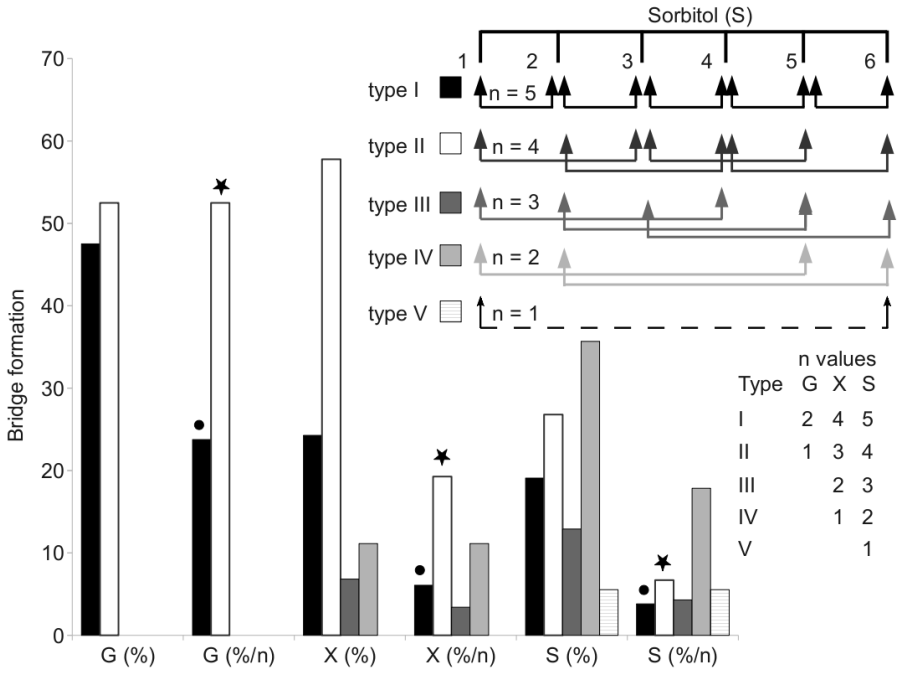

Fig. 24. Propensity of hydrogen bond bridge formation. (Inset) Sorbitiol molecule stick diagram with six -OH groups (one on each carbon atom). Gray-scale bonding pattern shown by arrows indicates the number of different variations or multiplicity (n) of hydrogen bond bridges for different bridge types. For example, type IV bridge between -OH groups separated by five carbon atoms can form in two ways (n=2). (a) Bar graph of percent bridge formation and percent averaged by multiplicity for each bridge type between alcohols and collagen (%/n). Circle and star indicate hydrogen bond bridges favors -OH groups in 1,3 over 1,2 positions. (b) Bond multiplicity (n) of glycerol (G), xylitol (X), and sorbitol (S).

Figure 24 shows the fraction of different bridge types along with the normalized values based on the multiplicity (n) of each bridge type. Depending on the separations of hydroxyl groups, certain bridge types have higher propensities to form. For all alcohols in the MD simulation, type II bridges were more favored than type I bridges (indicated by star and circle in Fig. 24). Xylitol and sorbitol, which have more than three hydroxyl groups, indicate a preference for type II and type IV bridges compared to type I or type III bridges. The difference between xylitol and sorbitol is their preference for type II and type IV bridges. While xylitol, like glycerol, prefers type II bridges (Fig. 24), sorbitol prefers to form type IV bridges. Although more tests are necessary to clarify the above behavior, it provides a plausible explanation for similar optical clearing properties of xylitol and glycerol. Type IV bridges of sorbitol, on the other hand, spans across the surface of the collagen triple helix, thereby screening collagen-water interactions much more effectively than xylitol.

D. Discussion

This study builds on previous results that showed skin optical clearing induced by sugars and sugar-alcohols correlated with collagen solubility [93, 83]. Optical clearing potential of these agents was shown to increase with molecular weight with an intriguing dependence on hydroxyl group position. In particular, 1,3-propanediol was shown to have twice the optical clearing potential of 1,2-propanediol even though they had identical molecular weights (76.10 Da), similar refractive index (1.44 versus 1.43) and osmolality (8.3 versus 8.7 Osm/kg) [93, 83]. The optical clearing potentials of 1,2- and 1,3-propanediol reflected those of analogous sugar-alcohols ethylene glycol and glycerol, respectively, and suggested that hydroxyl group position was an important factor in an agent's ability to induce skin clearing. Our simulation re-

sults clearly indicate that the position of hydroxyl groups in alcohols affects their bridging ability. (Note that 1,2-propanediol and ethylene glycol can only form type I bridges while 1,3-propanediol only forms type II bridges.) This suggestion was further supported by the result that xylitol (five carbon sugar-alcohol) and glycerol (three carbon sugar-alcohol) exhibited half the optical clearing potential of sorbitol (six carbon sugar-alcohol), demonstrating that optical clearing was not solely dependent on molecular weight. Xylitol, like glycerol, preferentially forms type II rather than type IV bridges. This result provides a possible molecular basis for understanding similar optical clearing properties of xylitol and glycerol even though their molecular sizes are different. On the other hand, sorbitol preferred to form type IV bridges which span across the collagen surface and likely disrupt hydration layers more effectively.

Optical clearing potential of the sugar-alcohols correlated with the preferred bridge type identified by our MD simulations, which is also consistent with what has been observed with nonlinear optical microscopy using second harmonic generation in collagenous tissues during tissue optical clearing [85, 93]. Interactions that organize and assemble collagen molecules are mediated by hydration forces involving water bridges [85]. Earlier studies suggest that water bridges stabilize collagen tertiary structures and that hydration shell organizes triple helices for higher order assembly [85, 11]. Characteristic of chemical agents with significant OCP is the ability to form hydrogen bond bridges which would disrupt hydration layers and affect interaction forces mediating collagen self-assembly.

E. Conclusion

MD simulations were used to elucidate and support experimental measurements on chemical agent interactions that drive (collagenous) tissue optical clearing. Tissue

optical clearing is driven by the disruption and replacement of collagen hydration layer with chemical agent. Hyperosmotic sugar and sugar-alcohol solutions will induce equivalent temporary reduction in light scattering of native and fixed rodent skin, given a long enough exposure time. However, the rate at which agent-induced optical clearing will occur depends on the surface hydrogen bond bridge formation. This rate is correlated with occupation of collagen hydrogen bonding sites, with higher optical clearing rates for agents with preference for hydrogen bond bridge formation with hydroxyl groups at specific positional separation.

CHAPTER IV

CONCLUSION

Water and hydration play a vital role in biomolecular surface interactions. These forces are often overlooked when modeling and studying biomolecular surface interactions, like for example drug design. Our study clearly shows that the magnitude of hydration forces can be comparable or in certain cases larger than protein-protein interactions and hence they cannot be overlooked. Our study can also be potentially used to improve current models that calculate biomolecular interaction energies by more accurately accounting for hydration forces.

The hydration map analysis scheme used is a novel method to understand the properties of hydration at high resolution as never seen before. This method can be easily applied to any biomolecular system of interest to study hydration water properties on the surface. The hydration maps show the ubiquitous nature of the primary hydration shell around protein molecules irrespective of the underlying aminoacid sequence and polarity. The primary hydration shell is formed irrespective of the different orientation and hydrogen bonding abilities of the water on the surface. This suggests that, though specific protein-water interactions could play a role, the boundary packing effect of water molecules play an important role in the formation of the primary hydration shell.

The oscilating nature of the hydration force, with ocilating wavelength similar to the diameter of a water molecule, suggests the structural origin of hydration force due to removal of a layer of water from between the surfaces. On the other hand, very little change in force magnitudes at different temperatures suggests that dynamics of water possibly plays very little role in force generation. We hypothesize that the temperature dependence of protein assembly and binding seen in experiments should

be because of change in protein motion and dynamics rather than that of water.

The qualitative features of measured force curves were similar in all the three systems suggesting that the similar oscilating nature of hydration forces can be expected in most biomolecular surface interactions. The magnitude of the forces depend on the hydration nature (dry or wet), geometry, and flexibility of the interacting surfaces and not just the polarity or amino acid sequence of the surfaces alone.

We also studied the hydrogen bonding of alcohol molecules on collagen surface which helped explain the dependence of position -OH bonds in alcohols on its clearing ability. By disrupting the hydration shell, alcohols would hinder the lubricating nature of hydration shells and hence the assembly of collagen molecules. Alcohol surface interactions would also reduce the protein-protein attractive interactions between collagen molecules there by causing optical clearing.

A major limitation, and a possible topic for future study, is the effect of ions on hydration and assembly. Many biomolecular interaction depend on the concentration of ions and it is unclear how they may enhance or hinder protein assembly and binding. Though from our optical clearing studies, we can hypothesize that any disruption of the hydration shell would affect the ability of biomolecules to assemble or bind, it is still unclear and hence studying effect of ions would be crucial to increase our understanding of protein assembly.

REFERENCES

[1] J. Myllyharju and K. I. Kivirikko, "Collagens and collagen-related diseases." *Annal Med*, vol. 33, no. 1, pp. 7–21, 2001.

[2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell.* New York: Garland Science, 2000.

[3] G. Veit, B. Kobbe, D. R. Keene, M. Paulsson, M. Koch, and R. Wagener, "Collagen XXVIII, a novel von Willebrand factor a domain-containing protein with many imperfections in the collagenous domain." *J Biol Chem*, vol. 281, no. 6, pp. 3494–3504, 2006.

[4] J. L. Lauer-Fields, D. Juska, and G. B. Fields, "Matrix metalloproteinases and collagen catabolism." *Biopolymers*, vol. 66, no. 1, pp. 19–32, 2002.

[5] G. N. Ramachandran and G. Karta, "Structure of collagen." *Nature*, vol. 176, no. 4482, pp. 593–595, 1955.

[6] R. Z. Kramer, J. Bella, P. Mayville, B. Brodsky, and H. M. Berman, "Sequence dependent conformational variations of collagen triple-helical structure." *Nat. Struct. Biol.*, vol. 6, no. 5, pp. 454–457, May 1999.

[7] L. Chung, D. Dinakarpandian, N. Yoshida, J. L. Lauer-Fields, G. B. Fields, R. Visse, and H. Nagase, "Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis." *EMBO J*, vol. 23, no. 15, pp. 3020–3030, 2004.

[8] A. V. Persikov and B. Brodsky, "Unstable molecules form stable tissues." *Proc Natl Acad Sci U S A*, vol. 99, no. 3, pp. 1101–1103, 2002.

[9] C. A. Miles and A. J. Bailey, "Studies of the collagen-like peptide (pro-pro-gly)$_{10}$ confirm that the shape and position of the type I collagen denaturation endotherm is governed by the rate of helix unfolding." *J Mol Biol*, vol. 337, no. 4, pp. 917–931, 2004.

[10] J. A. Hodges and R. T. Raines, "Stereoelectronic and steric effects in the collagen triple helix: toward a code for strand association." *J Am Chem Soc*, vol. 127, no. 45, pp. 15 923–15 932, 2005.

[11] K. M. Ravikumar and W. Hwang, "Region-specific role of water in collagen unwinding and assembly." *Proteins*, vol. 72, no. 4, pp. 1320–1332, Apr 2008.

[12] C. L. Brooks and M. Karplus, "Solvent effects on protein motion and protein effects on solvent motion. dynamics of the active site region of lysozyme." *J Mol Biol*, vol. 208, no. 1, pp. 159–181, 1989.

[13] P. Ball, "Water as an active constituent in cell biology." *Chem. Rev.*, vol. 108, no. 1, pp. 74–108, Jan 2008.

[14] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics." *J. Mol. Graphics.*, vol. 14, no. 1, pp. 33–38, 1996.

[15] G. B. Fields, "A model for interstitial collagen catabolism by mammalian collagenases." *J Theor Biol*, vol. 153, no. 4, pp. 585–602, 1991.

[16] K. E. Kadler, Y. Hojima, and D. J. Prockop, "Assembly of type I collagen fibrils de novo. Between 37 and 41 degrees C the process is limited by micro-unfolding of monomers." *J Biol Chem*, vol. 263, no. 21, pp. 10 517–10 523, 1988.

[17] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "A program for macromolecular energy, minimization, and dynamics

calculations," *J Comput Chem*, vol. 4, pp. 187–217, 1983.

[18] K. M. Ravikumar, J. D. Humphery, and W. Hwang, "Spontaneous unwinding of a labile domain in a collagen triple helix," *J Mech Mater Struct*, vol. 2, pp. 999–1010, 2007.

[19] R. Z. Kramer, L. Vitagliano, J. Bella, R. Berisio, L. Mazzarella, B. Brodsky, A. Zagari, and H. M. Berman, "X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly)." *J Mol Biol*, vol. 280, no. 4, pp. 623–638, Jul 1998.

[20] J. Israelachvili and H. Wennerström, "Role of hydration and water structure in biological and colloidal interactions." *Nature*, vol. 379, no. 6562, pp. 219–225, Jan 1996.

[21] Y. Levy and J. N. Onuchic, "Water mediation in protein folding and molecular recognition." *Annu Rev Biophys Biomol Struct*, vol. 35, pp. 389–415, 2006.

[22] C. Schröder, T. Rudas, S. Boresch, and O. Steinhauser, "Simulation studies of the protein-water interface. I. properties at the molecular resolution." *J Chem Phys*, vol. 124, no. 23, p. 234907, 2006.

[23] V. Makarov, B. M. Pettitt, and M. Feig, "Solvation and hydration of proteins and nucleic acids: a theoretical view of simulation and experiment." *Accounts Chem Res*, vol. 35, no. 6, pp. 376–384, 2002.

[24] K. Okuyama, C. Hongo, G. Wu, K. Mizuno, K. Noguchi, S. Ebisuzaki, Y. Tanaka, N. Nishino, and H. P. Bächinger, "High-resolution structures of collagen-like peptides [(Pro-Pro-Gly)$_4$-Xaa-Yaa-Gly-(Pro-Pro-Gly)$_4$]: Implications for triple-helix hydration and Hyp(X) puckering." *Biopolymers*, vol. 91, no. 5, pp. 361–372,

May 2009.

[25] R. Z. Kramer, J. Bella, B. Brodsky, and H. M. Berman, "The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence." *J Mol Biol*, vol. 311, no. 1, pp. 131–147, 2001.

[26] J. T. Nielsen, M. Bjerring, M. D. Jeppesen, R. O. Pedersen, J. M. Pedersen, K. L. Hein, T. Vosegaard, T. Skrydstrup, D. E. Otzen, and N. C. Nielsen, "Unique identification of supramolecular structures in amyloid fibrils by solid-state NMR spectroscopy." *Angew Chem Int Ed Engl*, vol. 48, no. 12, pp. 2118–2121, 2009.

[27] R. Nelson, M. R. Sawaya, M. Balbirnie, A. . Madsen, C. Riekel, R. Grothe, and D. Eisenberg, "Structure of the cross-beta spine of amyloid-like fibrils." *Nature*, vol. 435, no. 7043, pp. 773–778, 2005.

[28] D. Leckband and J. Israelachvili, "Intermolecular forces in biology." *Q. Rev. Biophys.*, vol. 34, no. 2, pp. 105–267, May 2001.

[29] M. Chaplin, "Do we underestimate the importance of water in cell biology?" *Nat Rev Mol Cell Biol*, vol. 7, no. 11, pp. 861–866, 2006.

[30] B. Halle, "Protein hydration dynamics in solution: a critical survey." *Philos. Trans. R. Soc. Lond. B*, vol. 359, no. 1448, pp. 1207–1223, Aug 2004.

[31] K. Lum, D. Chandler, and J. D. Weeks, "Hydrophobicity at small and large length scales," *J Phys Chem B*, vol. 103, no. 22, pp. 4570–4577, 1999.

[32] D. Chandler, "Interfaces and the driving force of hydrophobic assembly." *Nature*, vol. 437, no. 7059, pp. 640–647, 2005.

[33] B. J. Berne, J. D. Weeks, and R. Zhou, "Dewetting and hydrophobic interaction in physical and biological systems." *Annu Rev Phys Chem*, vol. 60, pp. 85–103, 2009.

[34] M. Manciu and E. Ruckenstein, "Oscillatory and monotonic polarization. The polarization contribution to the hydration force," *Langmuir*, vol. 17, no. 24, pp. 7582–7592, 2001.

[35] F. Bresme and A. Wynveen, "On the influence of solute polarizability on the hydrophobic interaction." *J Chem Phys*, vol. 126, no. 4, p. 044501, 2007.

[36] F. Despa and R. S. Berry, "The origin of long-range attraction between hydrophobes in water." *Biophys J*, vol. 92, no. 2, pp. 373–378, 2007.

[37] E. E. Meyer, K. J. Rosenberg, and J. Israelachvili, "Recent progress in understanding hydrophobic interactions." *Proc Natl Acad Sci U S A*, vol. 103, no. 43, pp. 15 739–15 746, 2006.

[38] J. Israelachvili, *Intermolecular and Surface Forces*, 3rd ed. London: Academic Press, 2010.

[39] C. Eun and M. L. Berkowitz, "Origin of the hydration force: water-mediated interaction between two hydrophilic plates." *J Phys Chem B*, vol. 113, no. 40, pp. 13 222–13 228, Oct 2009.

[40] R. Baron, P. Setny, and J. A. McCammon, "Water in cavity-ligand recognition." *J Am Chem Soc*, vol. 132, no. 34, pp. 12 091–12 097, 2010.

[41] N. Besseling, "Theory of hydration forces between surfaces," *Langmuir*, vol. 13, no. 7, pp. 2113–2122, 1997.

[42] L. Lu and M. L. Berkowitz, "Hydration force between model hydrophilic surfaces: computer simulations." *J Chem Phys*, vol. 124, no. 10, p. 101101, Mar 2006.

[43] N. Giovambattista, C. F. Lopez, P. J. Rossky, and P. G. Debenedetti, "Hydrophobicity of protein surfaces: Separating geometry from chemistry." *Proc Natl Acad Sci U S A*, vol. 105, no. 7, pp. 2274–2279, 2008.

[44] N. Choudhury and B. M. Pettitt, "On the mechanism of hydrophobic association of nanoscopic solutes." *J Am Chem Soc*, vol. 127, no. 10, pp. 3556–3567, 2005.

[45] R. Zhou, X. Huang, C. J. Margulis, and B. J. Berne, "Hydrophobic collapse in multidomain protein folding." *Science*, vol. 305, no. 5690, pp. 1605–1609, Sep 2004.

[46] L. Hua, X. Huang, R. Zhou, and B. J. Berne, "Dynamics of water confined in the interdomain region of a multidomain protein." *J Phys Chem B*, vol. 110, no. 8, pp. 3704–3711, Mar 2006.

[47] P. Liu, X. Huang, R. Zhou, and B. J. Berne, "Observation of a dewetting transition in the collapse of the melittin tetramer." *Nature*, vol. 437, no. 7055, pp. 159–162, Sep 2005.

[48] J. L. MacCallum, M. S. Moghaddam, H. S. Chan, and D. P. Tieleman, "Hydrophobic association of alpha-helices, steric dewetting, and enthalpic barriers to protein folding." *Proc Natl Acad Sci U S A*, vol. 104, no. 15, pp. 6206–6210, 2007.

[49] Y. K. Cheng and P. J. Rossky, "Surface topography dependence of biomolecular hydrophobic hydration." *Nature*, vol. 392, no. 6677, pp. 696–699, 1998.

[50] J. Mittal and G. Hummer, "Interfacial thermodynamics of confined water near molecularly rough surfaces." *Faraday Discuss*, vol. 146, pp. 341–52; discussion 367–93, 395–401, 2010.

[51] P. Setny, R. Baron, and J. A. McCammon, "How can hydrophobic association be enthalpy driven?" *J Chem Theory Comput*, vol. 6, no. 9, pp. 2866–2871, Sep 2010.

[52] J. Bella, B. Brodsky, and H. M. Berman, "Hydration structure of a collagen peptide." *Structure*, vol. 3, no. 9, pp. 893–906, 1995.

[53] S. Leikin, D. C. Rau, and V. A. Parsegian, "Temperature-favoured assembly of collagen is driven by hydrophilic not hydrophobic interactions." *Nat Struct Biol*, vol. 2, no. 3, pp. 205–210, Mar 1995.

[54] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. Ø. Madsen, C. Riekel, and D. Eisenberg, "Atomic structures of amyloid cross-beta spines reveal varied steric zippers." *Nature*, vol. 447, no. 7143, pp. 453–457, May 2007.

[55] F. Abraham, "The interfacial density profile of a Lennard-Jones fluid in contact with a (100) Lennard-Jones wall and its relationship to idealized fluid/wall systems: A Monte Carlo simulation," *J Chem Phys*, vol. 68, pp. 3713–3716, 1978.

[56] F. Merzel and J. C. Smith, "Is the first hydration shell of lysozyme of higher density than bulk water?" *Proc Natl Acad Sci U S A*, vol. 99, no. 8, pp. 5378–5383, 2002.

[57] D. V. D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C.

Berendsen, "GROMACS: Fast, flexible, and free." *J Comput Chem*, vol. 26, no. 16, pp. 1701–1718, Dec 2005.

[58] A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, "All-atom empirical potential for molecular modeling and dynamics studies of proteins." *J Phys Chem*, vol. 102, pp. 3586–3616, 1998.

[59] D. Anderson, "Collagen self-assembly: A complementary experimental and theoretical perspective." Ph.D. dissertation, University of Toronto, 2005.

[60] B. Hess, H. Bekker, H. Berendsen, and J. Fraaije, "LINCS: A linear constraint solver for molecular simulations," *J Comput Chem*, vol. 18, no. 12, pp. 1463–1472, 1997.

[61] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N log (N) method for Ewald sums in large systems," *J Chem Phys*, vol. 98, pp. 10 089–10 092, 1993.

[62] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling." *J Chem Phys*, vol. 126, no. 1, p. 014101, 2007.

[63] H. J. C. Berendsen, J. P. M. Postma, W. M. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *J Chem Phys*, vol. 81, pp. 3684–3690, 1984.

[64] P. Mark and L. Nilsson, "Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K," *J Phys Chem A*, vol. 105, no. 43, pp. 9954–9960, 2001.

[65] H. De Loof, L. Nilsson, and R. Rigler, "Molecular dynamics simulation of galanin in aqueous and nonaqueous solution," *J Am Chem Soc*, vol. 114, no. 11, pp. 4028–4035, 1992.

[66] J. P. R. O. Orgel, T. C. Irving, A. Miller, and T. J. Wess, "Microfibrillar structure of type I collagen in situ." *Proc Natl Acad Sci U S A*, vol. 103, no. 24, pp. 9001–9005, 2006.

[67] A. Bizzarri and S. Cannistraro, "Molecular dynamics of water at the protein-solvent interface," *J Phys Chem B*, vol. 106, no. 26, pp. 6617–6633, 2002.

[68] H. Kovacs, A. E. Mark, and W. F. van Gunsteren, "Solvent structure at a hydrophobic protein surface." *Proteins*, vol. 27, no. 3, pp. 395–404, 1997.

[69] D. Russo, G. Hura, and T. Head-Gordon, "Hydration dynamics near a model protein surface," *Biophys J*, vol. 86, no. 3, pp. 1852–1862, 2004.

[70] D. Russo, J. Ollivier, and J. Teixeira, "Water hydrogen bond analysis on hydrophilic and hydrophobic biomolecule sites," *Phys Chem Chem Phys*, vol. 10, no. 32, pp. 4968–4974, 2008.

[71] J. Park, B. Kahng, and W. Hwang, "Thermodynamic selection of steric zipper patterns in the amyloid cross-beta spine." *PLoS Comput Biol*, vol. 5, no. 9, p. e1000492, 2009.

[72] I. A. Nyrkova, A. N. Semenov, A. Aggeli, and N. Boden, "Fibril stability in solutions of twisted β-sheet peptides: a new kind of micellization in chiral systems," *Eur Phys J B*, vol. 17, no. 3, pp. 481–497, 2000.

[73] L. Esposito, C. Pedone, and L. Vitagliano, "Molecular dynamics analyses of cross-beta-spine steric zipper models: beta-sheet twisting and aggregation." *Proc Natl Acad Sci U S A*, vol. 103, no. 31, pp. 11 533–11 538, Aug 2006.

[74] A. Gubskaya and P. Kusalik, "The total molecular dipole moment for liquid water," *J Chem Phys*, vol. 117, pp. 5290–5302, 2002.

[75] S. Leikin, D. C. Rau, and V. A. Parsegian, "Direct measurement of forces between self-assembled proteins: temperature-dependent exponential forces between collagen triple helices." *Proc Natl Acad Sci U S A*, vol. 91, no. 1, pp. 276–280, 1994.

[76] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein." *J Mol Biol*, vol. 157, no. 1, pp. 105–132, May 1982.

[77] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, and F. G. Parak, "Slaving: Solvent fluctuations dominate protein dynamics and functions," *Proc Natl Acad Sci U S A*, vol. 99, no. 25, pp. 16 047–16 051, 2002.

[78] A. D. Simone, G. G. Dodson, C. S. Verma, A. Zagari, and F. Fraternali, "Prion and water: tight and dynamical hydration sites have a key role in structural stability." *Proc Natl Acad Sci U S A*, vol. 102, no. 21, pp. 7535–7540, 2005.

[79] C. Malardier-Jugroot, M. Johnson, and T. Head-Gordon, "Aqueous peptides as experimental models for hydration water dynamics near protein surfaces," *Phys Chem Chem Phys*, vol. 10, no. 32, pp. 4903–4908, 2008.

[80] V. V. Tuchin, *Optical Clearing of Tissues and Blood*. Bellingham, WA: SPIE, 2005.

[81] G. Vargas, E. K. Chan, J. K. Barton, H. G. Rylander, and A. J. Welch, "Use of an agent to reduce scattering in skin." *Lasers Surg Med*, vol. 24, no. 2, pp. 133–141, 1999.

[82] V. V. Tuchin, R. K. Wang, and A. T. Yeh, "Optical clearing of tissues and cells." *J Biomed Opt*, vol. 13, no. 2, p. 021101, 2008.

[83] J. Hirshburg, B. Choi, J. S. Nelson, and A. T. Yeh, "Collagen solubility correlates with skin optical clearing." *J Biomed Opt*, vol. 11, no. 4, p. 040501, 2006.

[84] L. Leonardi, A. Ruggeri, N. Roveri, A. Bigi, and E. Reale, "Light microscopy, electron microscopy, and x-ray diffraction analysis of glycerinated collagen fibers." *J Ultrastruct Res*, vol. 85, no. 2, pp. 228–237, 1983.

[85] A. T. Yeh, B. Choi, J. S. Nelson, and B. J. Tromberg, "Reversible dissociation of collagen in tissues." *J Invest Dermatol*, vol. 121, no. 6, pp. 1332–1335, 2003.

[86] N. Kuznetsova, S. L. Chi, and S. Leikin, "Sugars and polyols inhibit fibrillogenesis of type i collagen by disrupting hydrogen-bonded water bridges between the helices." *Biochemistry*, vol. 37, no. 34, pp. 11 888–11 895, 1998.

[87] J. M. Hirshburg, K. M. Ravikumar, W. Hwang, and A. T. Yeh, "Molecular basis for optical clearing of collagenous tissues." *J Biomed Opt*, vol. 15, no. 5, p. 055002, 2010.

[88] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M.

York, and M. Karplus, "CHARMM: The biomolecular simulation program." *J Comput Chem*, vol. 30, no. 10, pp. 1545–1614, 2009.

[89] W. Im, M. S. Lee, and C. L. Brooks, "Generalized born model with a simple smoothing function." *J Comput Chem*, vol. 24, no. 14, pp. 1691–1702, 2003.

[90] N. K. Shah, J. A. Ramshaw, A. Kirkpatrick, C. Shah, and B. Brodsky, "A host-guest set of triple-helical peptides: stability of Gly-X-Y triplets containing common nonpolar residues." *Biochemistry*, vol. 35, no. 32, pp. 10 262–10 268, 1996.

[91] J. K. Rainey and M. C. Goh, "An interactive triple-helical collagen builder." *Bioinformatics*, vol. 20, no. 15, pp. 2458–2459, 2004.

[92] A. T. Brünger and M. Karplus, "Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison." *Proteins*, vol. 4, no. 2, pp. 148–156, 1988.

[93] J. Hirshburg, B. Choi, J. S. Nelson, and A. T. Yeh, "Correlation between collagen solubility and skin optical clearing using sugars." *Lasers Surg Med*, vol. 39, no. 2, pp. 140–144, 2007.

APPENDIX A

HYDRATION MAP ANALYSIS

GROMACS code to analyze hydration.

```
static char *SRCID_template_c = "$Id: template.c,v 1.5 2008/05/29 08:36:53 hess Exp $";

#include <gromacs/statutil.h>
#include <gromacs/typedefs.h>
#include <gromacs/smalloc.h>
#include <gromacs/vec.h>
#include <gromacs/copyrite.h>
#include <gromacs/tpxio.h>

/* Places to edit in the file

1) MAX_W (20000) - maximum number of water molecule sin the system


2) Maximum Protein oxygen atoms = 5000
   Maximum protein hydrogen atoms = 5000

*/

const  int MAX_W=500000;        // EDIT: Maximum Water LIMIT
const  int MAX_PROTEIN=20000;   // EDIT: Maximum Protein Oxygen, Hydrogen atoms LIMIT

//To add two vectors
void add_vec(float a[], float b[], float c[])
  {
    c[0]= (a[0]+b[0]);
    c[1]= (a[1]+b[1]);
    c[2]= (a[2]+b[2]);
  }

//To subtract two vectors
void sub_vec(float a[], float b[], float c[])
  {
    c[0]= (a[0]-b[0]);
    c[1]= (a[1]-b[1]);
    c[2]= (a[2]-b[2]);
  }

// To get modulus of a vector
void mod_vec(float a[], float *ans)
  {
    float temp;
    temp = (a[0]*a[0])+(a[1]*a[1])+(a[2]*a[2]);
    *ans = sqrt(temp);
  }

// To get unit vector along a vector
void unit_vec(float a[], float unit[])
  {
    float temp;
    temp = (a[0]*a[0])+(a[1]*a[1])+(a[2]*a[2]);
    temp = sqrt(temp);
    unit[0] = a[0]/temp;
    unit[1] = a[1]/temp;
    unit[2] = a[2]/temp;
  }

//To find cross prod between two vectors
void cross_vec(float a[], float b[], float c[])
  {
    c[0]= (a[1]*b[2]-a[2]*b[1]);
    c[1]= (a[2]*b[0]-a[0]*b[2]);
    c[2]= (a[0]*b[1]-a[1]*b[0]);
  }

/* Put the coordinates of a[] in a
    defined bin and return the
    bin number as a coordinate */
void get_bin(float a[], int bin[], float X_BOX_SIZE, float Y_BOX_SIZE, float Z_BOX_SIZE)
  {
    int X_MAX_BINS=X_BOX_SIZE*10, Y_MAX_BINS=Y_BOX_SIZE*10, Z_MAX_BINS=Z_BOX_SIZE*10;
    int x_bin=0, y_bin=0, z_bin=0;
```

```
    x_bin = (a[0]/(X_BOX_SIZE/X_MAX_BINS));
    y_bin = (a[1]/(Y_BOX_SIZE/Y_MAX_BINS));
    z_bin = (a[2]/(Z_BOX_SIZE/Z_MAX_BINS));

    /* Due to pressure coupling the box size might
       increase slightly > *_BOX_SIZE:
       In that case make the bin as MAX_BINS-1 */
    if(x_bin >= X_MAX_BINS){x_bin = X_MAX_BINS-1;}
    if(y_bin >= Y_MAX_BINS){y_bin = Y_MAX_BINS-1;}
    if(z_bin >= Z_MAX_BINS){z_bin = Z_MAX_BINS-1;}

    bin[0] = x_bin;
    bin[1] = y_bin;
    bin[2] = z_bin;
  }

//To get dot product between two vectors
void dot_vec(float a[], float b[], float *ans)
  {
    float c[3];
    c[0]= (a[0]*b[0]);
    c[1]= (a[1]*b[1]);
    c[2]= (a[2]*b[2]);
    *ans = c[0]+c[1]+c[2];
  }

// To get the angle between vectors
// in degrees
void angle_bet_vec(float a[], float b[], float *theta)
  {
    float  dot_p, mod1, mod2,check;
    dot_vec(a,b,&dot_p);
    mod_vec(a,&mod1);
    mod_vec(b,&mod2);
    check = (dot_p/(mod1*mod2));
    if(check >=  1){printf("CAREFUL cos(theta) %f > 1!!",check); check =  0.999999;}
    if(check <= -1){printf("CAREFUL cos(theta) %f < 1!!",check); check = -0.999999;}
    *theta = acos(check)*180.0/3.1415;
  }

//Get H-bonding angle for ow and nei_h
float get_angle(float ow[], float nei_h[], float nei_o[])
  {
    float a[3],b[3],angle;
    a[0] = ow[0] - nei_h[0];
    a[1] = ow[1] - nei_h[1];
    a[2] = ow[2] - nei_h[2];

    b[0] = nei_h[0] - nei_o[0];
    b[1] = nei_h[1] - nei_o[1];
    b[2] = nei_h[2] - nei_o[2];

    angle_bet_vec(a, b, &angle);
    return angle;
  }

//Get distance between points t[], b[]
float get_distance(float t[], float b[], float X_BOX_SIZE, float Y_BOX_SIZE, float Z_BOX_SIZE)
  {
    int X_MAX_BINS=X_BOX_SIZE*10, Y_MAX_BINS=Y_BOX_SIZE*10, Z_MAX_BINS=Z_BOX_SIZE*10;
    float x_comp, y_comp, z_comp, dist;
    float a[3];
    a[0] = t[0]; a[1] = t[1]; a[2] = t[2];

    x_comp = fabs(a[0]-b[0]);
    y_comp = fabs(a[1]-b[1]);
    z_comp = fabs(a[2]-b[2]);

    if(fabs(x_comp) >= X_BOX_SIZE/2.0 && x_comp > 0 ){x_comp = X_BOX_SIZE - x_comp;}
    if(fabs(y_comp) >= Y_BOX_SIZE/2.0 && y_comp > 0 ){y_comp = Y_BOX_SIZE - y_comp;}
    if(fabs(z_comp) >= Z_BOX_SIZE/2.0 && z_comp > 0 ){z_comp = Z_BOX_SIZE - z_comp;}

    dist = (x_comp*x_comp + y_comp*y_comp + z_comp*z_comp);
    return sqrt(dist);
  }

//Get sq(distance) between vectors t[], b[]
float sq_distance_vec(float t[], float b[], float X_BOX_SIZE, float Y_BOX_SIZE, float Z_BOX_SIZE)
  {
    int X_MAX_BINS=X_BOX_SIZE*10, Y_MAX_BINS=Y_BOX_SIZE*10, Z_MAX_BINS=Z_BOX_SIZE*10;
    float x_comp, y_comp, z_comp, dist;
    float a[3];
    a[0] = t[0]; a[1] = t[1]; a[2] = t[2];

    x_comp = (a[0]-b[0]);
    y_comp = (a[1]-b[1]);
    z_comp = (a[2]-b[2]);
```

```
    if(fabs(x_comp) >= X_BOX_SIZE/2.0 && x_comp < 0 ){a[0] = a[0] + X_BOX_SIZE;}
    if(fabs(x_comp) >= X_BOX_SIZE/2.0 && x_comp > 0 ){a[0] = a[0] - X_BOX_SIZE;}

    if(fabs(y_comp) >= Y_BOX_SIZE/2.0 && y_comp < 0 ){a[1] = a[1] + Y_BOX_SIZE;}
    if(fabs(y_comp) >= Y_BOX_SIZE/2.0 && y_comp > 0 ){a[1] = a[1] - Y_BOX_SIZE;}

    if(fabs(z_comp) >= Z_BOX_SIZE/2.0 && z_comp < 0 ){a[2] = a[2] + Z_BOX_SIZE;}
    if(fabs(z_comp) >= Z_BOX_SIZE/2.0 && z_comp > 0 ){a[2] = a[2] - Z_BOX_SIZE;}

    x_comp = fabs(a[0] - b[0]);
    y_comp = fabs(a[1] - b[1]);
    z_comp = fabs(a[2] - b[2]);
    dist = (x_comp*x_comp + y_comp*y_comp + z_comp*z_comp);
    return dist;
  }



// Print float vector
void p_vec(float a[])
{
  printf("%f   %f   %f\n",a[0], a[1], a[2]);
}

// Print int vector
void i_vec(int a[])
{
  printf("%d   %d   %d\n",a[0], a[1], a[2]);
}


/* Main function */
int main(int argc,char *argv[])
{
  const char *desc[] = {
    "Water analysis code -Krishna(krishhere@mcbm.tamu.edu) 04/14/2010\n"
    " Hydration Map Analysis:\n\n"
  };

  /* Extra arguments - but note how you always get the begin/end
   * options when running the program, without mentioning them here!
   */

  gmx_bool bTop, b_density=TRUE, b_diff=FALSE, b_hbonds=FALSE;
  int        ePBC;
  char       title[STRLEN];
  t_topology top;
  t_trxframe fr;
  rvec       *xtop;
  matrix     topbox;
  t_trxstatus   *status;
  int        flags = TRX_READ_X;
  output_env_t          oenv;

  t_pargs pa[] = {
    { "-density", FALSE, etBOOL, {&b_density},
      "Output density in density.dat" },
    { "-diffusion", FALSE, etBOOL, {&b_diff},
      "Output translational/rotational diffusion coefficient t_diffusion.dat/r_diffusion.dat" },
    { "-hbonds", FALSE, etBOOL, {&b_hbonds},
      "Output hydrogen bonding hbonds.dat "}
  };


  t_filenm fnm[] = {
    { efTRX, "-f", NULL, ffREAD },      /* and this for the trajectory */
    { efTPS,  NULL,  NULL, ffREAD }   /* this is for the topology */
  };

#define NFILE asize(fnm)

  CopyRight(stderr,argv[0]);

  /* This is the routine responsible for adding default options,
   * calling the X/motif interface, etc. */
  parse_common_args(&argc,argv,PCA_CAN_TIME | PCA_CAN_VIEW,
    NFILE,fnm,asize(pa),pa,asize(desc),desc,0,NULL,&oenv);

  /* We don't need any topology information to write the coordinates,
   * but to show how it works we start by writing the name and
   * charge of the selected atom. It returns a boolean telling us
   * whether the topology was found and could be read
   */


  /* If none of the options are selected then exit */
```

```
  if(!(b_density || b_diff || b_hbonds))
    {
      printf(" ERROR: Choose atleast one of the options -density/diffusion/hbonds");
      exit(0);
    }

  /* Read topology file */
    bTop = read_tps_conf(ftp2fn(efTPS,NFILE,fnm),title,&top,&ePBC,\
 &xtop,NULL,topbox,TRUE);
    sfree(xtop);
    if (!bTop)
    {
        gmx_fatal(FARGS,"Need a run input file for option -mol, -cv or -cf");
    }

    /* Print simulation box dimensions */
    printf("\n\nDouble check box dimensions -- %f %f %f\n",\
   topbox[XX][XX],topbox[YY][YY],topbox[ZZ][ZZ] );


  float X_BOX_SIZE=topbox[XX][XX], Y_BOX_SIZE=topbox[YY][YY];
  float Z_BOX_SIZE=topbox[ZZ][ZZ];
  int X_MAX_BINS=X_BOX_SIZE*10, Y_MAX_BINS=Y_BOX_SIZE*10, Z_MAX_BINS=Z_BOX_SIZE*10;
  int starting_ow_index=0;
  int frame=0;
  int i,j,k,a,b,c,num_atoms,r;
  float mag_temp, mag_temp1, mag_temp2, mag_temp3;
  float **ow, **h1, **h2, **prev_ow, **owh1, **owh2;
  float **dipole, **p_dipole, **pp_dipole;
  float **unit_dipole, **unit_p_dipole, **unit_pp_dipole;
  float **prev_unit_dipole, **prev_unit_p_dipole, **prev_unit_pp_dipole;

  ow =      (float **) (malloc(MAX_W * sizeof(float *)));
  h1 =      (float **) (malloc(MAX_W * sizeof(float *)));
  h2 =      (float **) (malloc(MAX_W * sizeof(float *)));
  prev_ow = (float **) (malloc(MAX_W * sizeof(float *)));
  owh1 =    (float **) (malloc(MAX_W * sizeof(float *)));
  owh2 =    (float **) (malloc(MAX_W * sizeof(float *)));

  dipole =    (float **) (malloc(MAX_W * sizeof(float *)));
  p_dipole =    (float **) (malloc(MAX_W * sizeof(float *)));
  pp_dipole =    (float **) (malloc(MAX_W * sizeof(float *)));

  unit_dipole =    (float **) (malloc(MAX_W * sizeof(float *)));
  unit_p_dipole =    (float **) (malloc(MAX_W * sizeof(float *)));
  unit_pp_dipole =    (float **) (malloc(MAX_W * sizeof(float *)));

  prev_unit_dipole =    (float **) (malloc(MAX_W * sizeof(float *)));
  prev_unit_p_dipole =    (float **) (malloc(MAX_W * sizeof(float *)));
  prev_unit_pp_dipole =    (float **) (malloc(MAX_W * sizeof(float *)));
  for(i = 0; i < MAX_W; i++)
{
  ow[i] =       (float *) (malloc(3 * sizeof(float )));
  h1[i] =       (float *) (malloc(3 * sizeof(float )));
  h2[i] =       (float *) (malloc(3 * sizeof(float )));
  prev_ow[i] =  (float *) (malloc(3 * sizeof(float )));
  owh1[i] =     (float *) (malloc(3 * sizeof(float )));
  owh2[i] =     (float *) (malloc(3 * sizeof(float )));

  dipole[i] =       (float *) (malloc(3 * sizeof(float )));
  p_dipole[i] =     (float *) (malloc(3 * sizeof(float )));
  pp_dipole[i] =    (float *) (malloc(3 * sizeof(float )));
  unit_dipole[i] =       (float *) (malloc(3 * sizeof(float )));
  unit_p_dipole[i] =     (float *) (malloc(3 * sizeof(float )));
  unit_pp_dipole[i] =    (float *) (malloc(3 * sizeof(float )));
  prev_unit_dipole[i] =       (float *) (malloc(3 * sizeof(float )));
  prev_unit_p_dipole[i] =     (float *) (malloc(3 * sizeof(float )));
  prev_unit_pp_dipole[i] =    (float *) (malloc(3 * sizeof(float )));

}


  float theta;
  int bin[3];
  //float sum_orientation1[X_MAX_BINS][Y_MAX_BINS][Z_MAX_BINS];
  //float sum_orientation2[X_MAX_BINS][Y_MAX_BINS][Z_MAX_BINS];
  //float sum_orientation3[X_MAX_BINS][Y_MAX_BINS][Z_MAX_BINS];
  float angle[3];
  int l,m,p,e,f,g;
  int min_x,max_x,min_y,max_y,min_z,max_z;
  int water_num, h;
  float nei_h1[3],nei_h2[3];;
  float h_dist1,h_dist2, h_angle1, h_angle2;
  float vec1[3],vec2[3],vec3[3];
  float com1[3],com2[3],com3[3];
  float x_sum=0,y_sum=0,z_sum=0;
  int num_of_O=0, prot_O[MAX_PROTEIN];
```

```
  int num_of_H=0, prot_H[MAX_PROTEIN];
  float prot_O_coord[MAX_PROTEIN][3], prot_H_coord[MAX_PROTEIN][3];

  int ***sum_density;        //holds total density of each unit cell
  int ***o_occupancy;        //OW atom number in each cell at each frame
  float ***sum_diffusion;    //total diffusion in each cell
  float ***sum_dipole;       // rotational diffusion of water dipole unit vector
  float ***sum_p_dipole;     // vector perpendicular to dipole in the the water plane
  float ***sum_pp_dipole;    // unit vec perpendiculat to both above vectors
  int ***hydro_bond;         // total hbonding in the unit cell

  sum_density =    (int ***) (malloc(X_MAX_BINS * sizeof(int **)));
  o_occupancy =    (int ***) (malloc(X_MAX_BINS * sizeof(int **)));
  sum_diffusion =  (float ***) (malloc(X_MAX_BINS * sizeof(float **)));
  sum_dipole =     (float ***) (malloc(X_MAX_BINS * sizeof(float **)));
  sum_p_dipole =   (float ***) (malloc(X_MAX_BINS * sizeof(float **)));
  sum_pp_dipole =  (float ***) (malloc(X_MAX_BINS * sizeof(float **)));
  hydro_bond =     (int ***) (malloc(X_MAX_BINS * sizeof(int **)));

  if(hydro_bond == NULL)
{
fprintf(stderr, "out of memory\n");
}
  for(i = 0; i < X_MAX_BINS; i++)
{
sum_density[i] =   (int **) (malloc(Y_MAX_BINS * sizeof(int *)));
o_occupancy[i] =   (int **) (malloc(Y_MAX_BINS * sizeof(int *)));
sum_diffusion[i] = (float **) (malloc(Y_MAX_BINS * sizeof(float *)));
sum_dipole[i] =    (float **) (malloc(Y_MAX_BINS * sizeof(float *)));
sum_p_dipole[i] =  (float **) (malloc(Y_MAX_BINS * sizeof(float *)));
sum_pp_dipole[i] = (float **) (malloc(Y_MAX_BINS * sizeof(float *)));
hydro_bond[i] =    (int **) (malloc(Y_MAX_BINS * sizeof(int *)));
if(hydro_bond[i] == NULL)
  {
    fprintf(stderr, "out of memory\n");
  }
for(j = 0; j < Y_MAX_BINS; j++)
  {
    sum_density[i][j] =   (int *) (malloc(Z_MAX_BINS * sizeof(int)));
    o_occupancy[i][j] =   (int *) (malloc(Z_MAX_BINS * sizeof(int)));
    sum_diffusion[i][j] = (float *) (malloc(Z_MAX_BINS * sizeof(float)));
    sum_dipole[i][j] =    (float *) (malloc(Z_MAX_BINS * sizeof(float)));
    sum_p_dipole[i][j] =  (float *) (malloc(Z_MAX_BINS * sizeof(float)));
    sum_pp_dipole[i][j] = (float *) (malloc(Z_MAX_BINS * sizeof(float)));
    hydro_bond[i][j] =    (int *) (malloc(Z_MAX_BINS * sizeof(int)));
    if(hydro_bond[i][j] == NULL)
      {
fprintf(stderr, "out of memory\n");
      }

    /* initialize arrays to 0 */
    for(k=0;k<Z_MAX_BINS;k++)
      {
sum_density[i][j][k] = 0;
sum_diffusion[i][j][k] = 0.0;
sum_dipole[i][j][k] = 0.0;
sum_p_dipole[i][j][k] = 0.0;
sum_pp_dipole[i][j][k] = 0.0;
//sum_orientation1[i][j][k] = 0.0;
//sum_orientation2[i][j][k] = 0.0;
//sum_orientation3[i][j][k] = 0.0;
o_occupancy[i][j][k] = -1;
//prev_o_occupancy[i][j][k] = -1;
hydro_bond[i][j][k] = 0;
      }

  }
}


  /* The first time we read data is a little special */
  read_first_frame(oenv,&status,ftp2fn(efTRX,NFILE,fnm),&fr,flags);


  /* set frame number = 0*/
  frame = 0;

  /* This is the main loop over frames */
    do {
    /* coordinates are available in the vector fr.x
     * you can find this and all other structures in
‘    * the types directory under the gromacs include dir.
     * Note how flags determines wheter to read x/v/f!
     */

      /* initialize cell occupancy arrays to -1 for each iteration */
      for(i=0;i<X_MAX_BINS;i++)
```

```
{
  for(j=0;j<Y_MAX_BINS;j++)
    {
      for(k=0;k<Z_MAX_BINS;k++)
{
  o_occupancy[i][j][k] = -1;
}
    }
}

      if(frame==0)
{
/* Store Protein Oxygen,Hydrogen coordinates for H-bond distance checking */
// Number of oxygens and hydrogens
// Initialize to 0
  num_of_O = 0;
  num_of_H = 0;
  /* Stop the loop when you encounter a water OW atom (or) total number of atoms */
  for(i=0; i<(top.atoms.nr) && !(strcmp(*top.atoms.atomname[i],"OW")==0); i++)
    {
      if(strcmp(*top.atoms.atomname[i],"O")==0 \
 || strcmp(*top.atoms.atomname[i],"OG2")==0 )
{
  prot_O[num_of_O] = i;
  num_of_O += 1;
}
      if(strcmp(*top.atoms.atomname[i],"HN")==0 \
 || strcmp(*top.atoms.atomname[i],"HG2")==0 )
{
  prot_H[num_of_H] = i;
  num_of_H += 1;
}
    }
  /* Print number of atoms considered for hbonding analysis */
  if(b_hbonds)
    {
      printf("\n\nFor hydrogen bonding analysis: \n");
      printf("Number of protein Oxygen atoms considered = %d \n", num_of_O);
      printf("Number of protein Hydrogens atoms considered= %d \n", num_of_H);
    }

  /* Store first atom index of OW */
  starting_ow_index = i;

  /* If thre are no water molecules in the system - exit */
  if(starting_ow_index+1 == (top.atoms.nr))
    {
      printf("THERE ARE NO WATER ATOMS IN THE SYSTEM\n");
      printf("EXITING\n");
      exit(0);
    }
  printf("\n\nCHECK - Starting water oxygen index seen  = %d\n",starting_ow_index+1);
  printf("Double check your configuration file to make sure...\n\n");
}


    /* assign bin numbers to Proteins Oxygen and Hydrogen atoms */
    if(frame>=0)
      {
// prot_O_coord[r] has coordinates of protein O atom r
for(r=0;r<num_of_O;r++)
  {
    prot_O_coord[r][0] = fr.x[prot_O[r]][XX];
    prot_O_coord[r][1] = fr.x[prot_O[r]][YY];
    prot_O_coord[r][2] = fr.x[prot_O[r]][ZZ];
  }
for(r=0;r<num_of_H;r++)
  {
    prot_H_coord[r][0] = fr.x[prot_H[r]][XX];
    prot_H_coord[r][1] = fr.x[prot_H[r]][YY];
    prot_H_coord[r][2] = fr.x[prot_H[r]][ZZ];
  }
      }

    /* For all water oxygen atoms */
    j = 0; // j is the total number of water oxygens

    // Starting atom number of water oxygen-1 (C array index starts at 0)
    // increment by 3 for 3 water atoms (ow,h1,h2)
    for(i=starting_ow_index; i<(top.atoms.nr) && (strcmp(*top.atoms.atomname[i],"OW")==0); i=i+3)
      {

if(b_density || b_diff || b_hbonds)
  {
    // assign coordinates of j th oxygen
    // and hydrogen atoms to ow[j], h1[j], h2[j]
    ow[j][0] = fr.x[i][XX];
```

```
    ow[j][1] = fr.x[i][YY];
    ow[j][2] = fr.x[i][ZZ];

    h1[j][0] = fr.x[i+1][XX];
    h1[j][1] = fr.x[i+1][YY];
    h1[j][2] = fr.x[i+1][ZZ];

    h2[j][0] = fr.x[i+2][XX];
    h2[j][1] = fr.x[i+2][YY];
    h2[j][2] = fr.x[i+2][ZZ];

    // get water oxygen bin number
    get_bin(ow[j], bin, X_BOX_SIZE, Y_BOX_SIZE, Z_BOX_SIZE);
    a = bin[0];
    b = bin[1];
    c = bin[2];

    // occupancy of bin a,b,c is the water number (j) or -1
    o_occupancy[a][b][c] = j;

    // increase density count of bin a,b,c by 1
    sum_density[a][b][c] += 1;
  }

if(b_diff)
  {
// Calculate unit vectors along
// 1. (r1) dipole
// 2. (r2) perpendicular to dipole (p_dipole)
// 3. (r3) perpendicular to both the above (pp_dipole)
sub_vec(h1[j],ow[j],owh1[j]);
sub_vec(h2[j],ow[j],owh2[j]);
add_vec(owh1[j],owh2[j],dipole[j]);
sub_vec(owh1[j],owh2[j],p_dipole[j]);

unit_vec(dipole[j],unit_dipole[j]);
unit_vec(p_dipole[j],unit_p_dipole[j]);
cross_vec(unit_dipole[j],unit_p_dipole[j],unit_pp_dipole[j]);


/* From the second frame onwards, compare r1,r2,r3 vectors
// for each water from the two consecutive frames and find their
// rotation angles.
// Sum rotation angles of each vector to the previous bin
// containing that water oxygen atom. */
if(frame>=1)
  {
    mag_temp = sq_distance_vec(ow[j],prev_ow[j], X_BOX_SIZE, Y_BOX_SIZE, Z_BOX_SIZE);

    dot_vec(unit_dipole[j],prev_unit_dipole[j], &mag_temp1);
    if(mag_temp1 >  1.00){mag_temp1 =  0.999999;}
    if(mag_temp1 < -1.00){mag_temp1 = -0.999999;}
    mag_temp1 = acos(mag_temp1);

    dot_vec(unit_p_dipole[j],prev_unit_p_dipole[j], &mag_temp2);
    if(mag_temp2 >  1.00){mag_temp2 =  0.999999;}
    if(mag_temp2 < -1.00){mag_temp2 = -0.999999;}
    mag_temp2 = acos(mag_temp2);

    dot_vec(unit_pp_dipole[j],prev_unit_pp_dipole[j], &mag_temp3);
    if(mag_temp3 >  1.00){mag_temp3 =  0.999999;}
    if(mag_temp3 < -1.00){mag_temp3 = -0.999999;}
    mag_temp3 = acos(mag_temp3);

    sum_diffusion[a][b][c] += mag_temp;
    sum_dipole[a][b][c]    += mag_temp1*mag_temp1;
    sum_p_dipole[a][b][c]  += mag_temp2*mag_temp2;
    sum_pp_dipole[a][b][c] += mag_temp3*mag_temp3;

  }
  }

// increment water oxygen
j++;
      }


    /* Again for all water atoms -- Second loop
       Variable j has the total number of oxygen atoms */
    // Hydogen bonds are calculated in the current frame

    for(k=0;k<j;k++)
      {

if(b_density || b_diff || b_hbonds)
  {
    // prev_* arrays store data/coordinates from the
```

```
    // previous frame
    prev_ow[k][0] = ow[k][0];
    prev_ow[k][1] = ow[k][1];
    prev_ow[k][2] = ow[k][2];
  }

if(b_diff)
  {
    prev_unit_dipole[k][0] = unit_dipole[k][0];
    prev_unit_dipole[k][1] = unit_dipole[k][1];
    prev_unit_dipole[k][2] = unit_dipole[k][2];
    prev_unit_p_dipole[k][0] = unit_p_dipole[k][0];
    prev_unit_p_dipole[k][1] = unit_p_dipole[k][1];
    prev_unit_p_dipole[k][2] = unit_p_dipole[k][2];
    prev_unit_pp_dipole[k][0] = unit_pp_dipole[k][0];
    prev_unit_pp_dipole[k][1] = unit_pp_dipole[k][1];
    prev_unit_pp_dipole[k][2] = unit_pp_dipole[k][2];
  }

if(b_hbonds)
  {
    // a,b,c has bin numbers of water oxygen k
    get_bin(ow[k], bin, X_BOX_SIZE, Y_BOX_SIZE, Z_BOX_SIZE);
    a = bin[0];
    b = bin[1];
    c = bin[2];

    // search for next oxygen atom
    // in a cube surrounding the current cell
    // NOTE: search for water oxygens not hydrogens
    // So the the cube has to be bigger (+5,-5)
    min_x = a-5; max_x = a+5;
    min_y = b-5; max_y = b+5;
    min_z = c-5; max_z = c+5;

    /* Count hydrogen bonding with Protein atoms*/
    // For each protein oxygen atom fing distance between
    // hydrogens of water k.
    // If distance <0.24 then increment H-bonding of
    // bin a,b,c by 1
    for(l=0;l<num_of_O;l++)
      {
h_dist1 = get_distance(h1[k],prot_O_coord[l], X_BOX_SIZE, Y_BOX_SIZE, Z_BOX_SIZE);
h_dist2 = get_distance(h2[k],prot_O_coord[l], X_BOX_SIZE, Y_BOX_SIZE, Z_BOX_SIZE);
if(h_dist1 < 0.24 || h_dist2 < 0.24)
  {
    hydro_bond[a][b][c] += 1;
  }
      }

    // Do the same as above for protein hydrogen atoms
    for(l=0;l<num_of_H;l++)
      {
h_dist1 = get_distance(ow[k],prot_H_coord[l], X_BOX_SIZE, Y_BOX_SIZE, Z_BOX_SIZE);
if(h_dist1 < 0.24 )
  {
    hydro_bond[a][b][c] += 1;
  }
      }

    // for each cell loop through the neighbouring cells
    // and check for hydrogen bonds.
    // If the cells are at the edges, then use
    // periodic boundary condition to find the neighbours
    // NOTE: Using periodic boundary condition is not accurate
    // when box size fluctuates under pressure coupling
    for(l=min_x;l<=max_x;l++)
      {
e = l;
if(l < 0){ e = e + X_MAX_BINS; }
if(l >= X_MAX_BINS){ e = e - X_MAX_BINS; }
for(m=min_y;m<=max_y;m++)
  {
    f = m;
    if(m < 0){f = f + Y_MAX_BINS;}
    if(m >= Y_MAX_BINS){f = f - Y_MAX_BINS;}
    for(p=min_z;p<=max_z;p++)
      {
g = p;
if(p < 0){g = g + Z_MAX_BINS;}
if(p >= Z_MAX_BINS){g = g - Z_MAX_BINS;}
if(o_occupancy[e][f][g] != -1 )
  {
    water_num = o_occupancy[e][f][g];
    nei_h1[0] = h1[water_num][0];
    nei_h1[1] = h1[water_num][1];
    nei_h1[2] = h1[water_num][2];
```

```
    nei_h2[0] = h2[water_num][0];
    nei_h2[1] = h2[water_num][1];
    nei_h2[2] = h2[water_num][2];
    h_dist1 = get_distance(ow[k],nei_h1, X_BOX_SIZE, Y_BOX_SIZE, Z_BOX_SIZE);
    h_dist2 = get_distance(ow[k],nei_h2, X_BOX_SIZE, Y_BOX_SIZE, Z_BOX_SIZE);
    //h_angle1 = get_angle(ow[k],nei_h1,nei_o);
    //h_angle2 = get_angle(ow[k],nei_h2,nei_o);
    // NOTE: Hydrogen bond angles are not checked.
    h_angle1=0.0;
    h_angle2=0.0;

    if(h_dist1 <= 0.24 && h_angle1 <=40.0 && k!= water_num )
      {
hydro_bond[a][b][c] += 1;
hydro_bond[e][f][g] += 1;
      }
    if(h_dist2 <= 0.24 && h_angle2 <=40.0 && k!= water_num )
      {
hydro_bond[a][b][c] += 1;
hydro_bond[e][f][g] += 1;
      }
  }
      }
  }
      }
  }
      }

    frame += 1;
    //if(fr.time>=2000.0){break;}
    } while(read_next_frame(oenv,status,&fr));   /* This is the main loop over frames */


  /* Print total water in the system as a check */
  printf("\nDouble check again... \nTotal number of water molecules analyzed = %d \n",j);
  printf("Total coordinate frames analyzed = %d \n",frame);


  /* Write Data to File */
  FILE *fpw1, *fpw2, *fpw3, *fpw4;

  if(b_density)
    {fpw1 = fopen("density.dat","w");}
  if(b_diff)
    {
      fpw2 = fopen("t_diffusion.dat","w");
      fpw3 = fopen("r_diffusion.dat","w");
    }
  if(b_hbonds)
    {fpw4 = fopen("hbonds.dat","w");}


  for(i=0;i<X_MAX_BINS;i++)
    {
      for(j=0;j<Y_MAX_BINS;j++)
{
  for(k=0;k<Z_MAX_BINS;k++)
    {
      if(b_density)
{
  fprintf(fpw1,"%4d %4d %4d %5d \n",i,j,k,sum_density[i][j][k]);
}

      if(b_diff)
{
  fprintf(fpw2,"%4d %4d %4d %5d %11.5f \n", i,j,k, \
  sum_density[i][j][k], sum_diffusion[i][j][k]);

  fprintf(fpw3,"%4d %4d %4d %5d %11.5f %11.5f %11.5f \n", i,j,k, \
  sum_density[i][j][k],sum_dipole[i][j][k], \
  sum_p_dipole[i][j][k], sum_pp_dipole[i][j][k]);
}

      if(b_hbonds)
{
  fprintf(fpw4,"%4d %4d %4d %5d %6d \n", i,j,k,\
 sum_density[i][j][k],hydro_bond[i][j][k]);
}

    }
}
    }

  fclose(fpw1);
  return 0;
}
```

VITA

Name: Krishnakumar Mayuram Ravikumar

Address:

337 Zachry Engineering Center

College Station, TX 77843-3120

krishhere@mcbm.tamu.edu

979-204-5014

Education:

M.S. Biomedical Engineering — May 2008

Texas A&M University

B.E. Mechanical Engineering — May 2004

University of Madras, India

Publications:

(1) K. M. Ravikumar and W. Hwang, "Role of hydration shell in the self-assembly of collagens and amyloid steric zipper filaments." (JACS In Print).

(2) J. M. Hirshburg, K. M. Ravikumar, W. Hwang, and A. T. Yeh, "Molecular basis for optical clearing of collagenous tissues." *J Biomed Opt*, vol. 15, no. 5, p. 055002, 2010.

(3) K. M. Ravikumar and W. Hwang, "Region-specific role of water in collagen unwinding and assembly." *Proteins*, vol. 72, no. 4, pp. 1320–1332, Apr 2008.

(4) K. M. Ravikumar, J. D. Humphery, and W. Hwang, "Spontaneous unwinding of a labile domain in a collagen triple helix," *J Mech Mater Struct*, vol. 2, pp. 999–1010, 2007.